Mechanistic and Evolutionary Analyses of the Sialic Acid Synthase Family

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biochemistry

at the

University of Canterbury

by

Dmitri Daniel Alexander Joseph

January 2014

Acknowledgements

It's hard to believe that it's finally over. It seems a ridiculous thing to say, but it really is hard to believe. When I started this work I was warned that finishing a PhD is a measure of willpower. It isn't just that the work is going to be difficult, but that working is going to be difficult. The sage advice of my primary supervisor was that a PhD is an emotional rollercoaster; there will be plenty of highs and lows. A major part of my PhD would constitute overcoming these hurdles; to focus on the triumphs and to learn to cut my losses. Truth be told I didn't really believe those words of wisdom at the time either…

What I can say without a doubt is that I could not have gotten this far without a lot of help from a lot of people. Firstly, my thanks must go to my primary supervisor, Professor Emily Parker, whose wisdom and advice have been essential during the course of my PhD. I have thoroughly enjoyed our lengthy discussions on all things science. Without your enthusiasm and optimism this PhD would have been a far harder task. Thank you for being such a cheerful, patient and approachable mentor.

I would like to thank the numerous members of the Parker group, both past and present for the friendships and fun times we have shared over the years. To Tammie Cookson, I don't think there are sufficient words to express how valuable our relationship is to me. I am incredibly lucky to have someone who I can trust, depend upon and who keeps me honest. You have made this PhD far more interesting and enjoyable than I could have possibly imagined. Thank you for being there for me both in and out of the lab.

To Sarah Wilson-Coutts, my first friend in the Chemistry department! Thank you for being such a kind and fantastic person. Your friendship will always be important to me. I have enjoyed our many talks about life and Lego. Thank you especially for liking my cake best. To Frances Huisman, thank you for your friendship. Thank you also for introducing me to Buffy, Better Off Ted, Firefly, and all sorts of interesting and nerdy series and movies that have kept me busy after work. To Michael 1.0, thank you…I think. You were a great help when 'discussing' science, thank you also for keeping me company whilst driving down to Queenstown.

Special thanks must go to the sialic acid synthase family. To my padawan Ryu Toyama, thank you for putting up with my humour, thank you for making me copious amounts of rManNAc and thank you for being a pleasure to work with. To my West Coast side-kick Thomas Cotton, thank you for all of your help in the lab. Thank you for all of the text messages and thank you for making me laugh. I am lucky to have had two such fantastic people to work alongside with.

To Gerd Mittelstädt thank you for our lengthy discussions on PCR, the Gateway® cloning system and computer games. To Gert-Jan Moggré thank you for our chats and your exquisite home brews. To Dr Wanting Jiao thank you for all of your crystallographic and modelling expertise. Thank you also for introducing me to Full Metal Alchemist. To Dr Richard Hutton, thank you for all of your help, the interesting conversations and the lunch-time walks. To Davey Lim, thank you for your invaluable friendship, guitar tutorials and superhero conversations. To my other science friends - Evan Nimmo, Nicky Blackmore, Eric Lang and everyone else, thank you for your company and friendship. To all of the members of the Parker group who have spent sleepless nights collecting crystallographic and SAXS data for me, thank you so much.

Thank you to my co-supervisor Professor Vic Arcus from the University of Waikato. Thank you for your crystallographic expertise and help in elucidating the *Ssu*SAS crystal structure. Thank you to both you and your group for putting up with me as an earthquake refugee! Thank you to both the Maurice-Wilkins Centre and College of Science for providing me with generous financial support. I would also like to acknowledge and thank the Maurice-Wilkins Centre for funding trips to Auckland for both conferences and networking. I would especially like to acknowledge Rochelle Ramsay for organising the trips and always making the Canterbury crew feel welcome.

I would like to acknowledge and thank my friends from outside the university environment. Thank you for keeping me sane and being there for me when I have needed you. I consider myself to be truly blessed to have so many fantastic people in my life. Thank you especially to Simon Hart, Matthew Baker, Kim Baker, David Farmer and Susan Farmer for your company, friendship, faith and all the happy memories. Thank you for reminding me what is important in life.

Finally to my family, who have always believed in me even when I have not. Thank you for your constant love and support. I cannot begin to describe how much each of you means to me. Thank you to my brother for your friendship and technical advice. Thank you to my grandparents and my aunts for all of your encouragement. I would especially like to acknowledge and thank my mum for always being there for me and for being so wonderful. Thank you to both you and Dad for encouraging my passion for science from childhood. I'd say it paid off pretty well.

Abstract

Sialic acids are prevalent in many organisms and facilitate a range of cellular processes in both bacteria and mammals. *N*-Acetylneuraminic acid (NANA) is the most common naturally occurring sialic acid and plays a key role in the pathogenesis of a select number of neuroinvasive bacteria such as *Neisseria meningitidis*. These pathogens coat themselves with polysialic acids, mimicking the exterior surface of mammalian cells and consequentially concealing the bacteria from the host's immune system. NANA is synthesised in prokaryotes via a condensation reaction between phosphoenolpyruvate (PEP) and *N*-acetylmannosamine (ManNAc). This reaction is catalysed by the domain swapped, homodimeric enzyme, *N*-acetylneuraminic acid synthase (NANAS). NANAS comprises two distinct domains; an N-terminal catalytic $(\beta/\alpha)_8$ barrel linked to a C-terminal antifreeze protein-like (AFPL) domain. The AFPL domain contributes a highly conserved arginine (Arg314) into the active site of the opposing monomer chain.

Chapter 2 of this thesis details the examination of Arg314 in *N. meningitidis* NANAS (*Nme*NANAS) through mutagenesis and a combination of kinetic and structural analyses. Using isothermal titration calorimetry and molecular modelling, Arg314 is found to be essential to the catalytic function of *Nme*NANAS. The delocalised, positively charged guanidinium functionality of Arg314 steers the sugar substrate ManNAc into a suitable position for reaction with PEP.

Chapter 3 details an investigation into the role of the AFPL domain by characterising a truncated variant of *Nme*NANAS. The results from this chapter demonstrate that the AFPL domain plays a critical role in both the catalytic function and quaternary structure stability of *Nme*NANAS. Molecular dynamics simulations and amino acid substitutions expose a complex hydrogen bonding relay, which links the roles of the catalytic and AFPL domains across subunit boundaries.

The work in Chapter 4 of this thesis examines the evolutionary relationships within the bacterial sialic acid synthase (SAS) family. An alignment of bacterial sequences identifies four distinct enzyme clades denoted as NANASs, legionaminic acid synthases (LegSs), pseudaminic acid synthases (PseSs) and a fourth clade with unknown catalytic activity termed Type IV SASs. The NANAS, LegS and PseS paralogues from *Campylobacter jejuni* as well as the Type IV SAS from *Chlorobium tepidum* were analysed to compare the relationships between these enzymes. Structural analysis techniques such as molecular modelling, substrate docking, small-angle X-ray scattering (SAXS) and analytical ultracentrifugation (AUC) were used to identify key similarities and differences between the four SAS homologues.

Chapter 5 of this thesis outlines the structural and kinetic analyses of the SAS from *Streptococcus suis*. The kinetic parameters of *Ssu*SAS were successfully obtained, revealing NANAS activity even though the enzyme was identified to be more closely related to members of the LegS clade. A crystal structure of *Ssu*SAS was also obtained at a resolution of 2.30 Å. Docking analyses using the crystal structure of *Ssu*SAS revealed favourable binding poses for the substrate of Leg biosynthesis, indicating that *Ssu*SAS may have LegS activity.

Chapter 6 details the cloning of two chimeric constructs. The catalytic and AFPL domains of *Nme*NANAS and *Homo sapiens N*-acetylneuraminic-9-phosphate synthase (NANA-9-PS) were interchanged in an effort to switch the activity of the two SAS enzymes. The *Nme*NANAS catalytic domain (*Nme*/*Hsa*) chimera was solubly expressed, however, the *Hsa*NANA-9-PS catalytic domain (*Hsa*/*Nme*) chimera was not. Although soluble, the *Nme*/*Hsa* chimera was not active as either a NANAS or NANA-9-PS. The *Nme*/*Hsa* chimera was found to have a destabilised quaternary state through AUC analysis, which is the likely cause of activity loss.

Chapter 7 discusses and interprets the results outlined in Chapters 2 to 6 and outlines the future prospects of the work begun in this thesis.

Table of contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 6

Chapter 7

Chapter 8

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 8

List of Abbreviations

A5P	D-arabinose 5-phosphate
AFPL	antifreeze protein-like
apo	*apo*enzyme
ATP	adenosine triphosphate
AUC	analytical ultracentrifugation
Bp	base pairs
CD	circular dichroism
Cje	*Campylobacter jejuni*
CMP	cytosine monophosphate
Co	cobalt
Cte	*Chlorobium tepidum*
CTP	cytosine triphosphate
DAH7PS	3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase
DNA	deoxyribose nucleic acid
dNTP	deoxynucleotide triphosphate
DSF	differential scanning fluorimetry
E4P	erythrose 4-phosphate
Eco	*Escherichia coli*
EPSPS	5-enolpyruvylshikimate-3-phosphate synthase
GalNAc	*N*-acetylgalactosamine
GlucNac	*N*-acetylglucosamine
Has	*Homo sapiens*
IMAC	immobilised metal affinity column
k_{cat}	turnover number
kDa	kilodalton

KDN	2-keto-3-deoxy-nonulosonic acid
KDN-9-P	2-keto-3-deoxy-nonulosonic acid-9-phosphate
KDN-9-PS	2-keto-3-deoxy-nonulosonic acid-9-phosphate synthase
KDO8PS	2-keto-3-deoxy-D-*manno*-octulosonate-8-phosphate synthase
K_i	inhibition constant
K_M	Michaelis-Menten constant
LB	lysogeny broth
Leg	legionaminic acid
Leg5Am7Ac	5-acetamidino-7-acetamido-3,5,7,9,-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid
Leg5AmNMe7Ac	5-*E*/*Z*-*N*-(*N*-methylacetimidoyl)-7-acetamidino-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid
LegS	legionaminic acid synthase
Lpn	*Legionella pneumophila*
ManNAc	*N*-acetylmannosamine
Mmu	*Mus musculus*
Mn	manganese
MurA	UDP-*N*-acetylglucosamine enolpyruvyl transferase
MWCO	molecular weight cut-off
NANA	*N*-acetylneuraminic acid
NANA-9-P	*N*-acetylneuraminic acid-9-phosphate
NANA-9-PS	*N*-acetylneuraminic acid-9-phosphate synthase
NANAS	*N*-acetylneuraminic acid synthase
NGNA	*N*-glycolylneuraminic acid
Nme	*Neisseria meningitidis*
NMR	nuclear magnetic resonance
ORF	open reading frame
PCR	polymerase chain reaction

PDB	Protein Data Bank
PEP	phosphoenolpyruvate
P_i	inorganic phosphate
PP_i	inorganic pyrophosphate
Pse	pseudaminic acid
Pse5Ac7Am	5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid
Pse5Am5Ac	5-acetamido-7-acetamidino-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid
PseS	pseudaminic acid synthase
RMSD	root mean square deviation
Sag	*Streptococcus agalactiae*
SAS	sialic acid synthase
SAXS	small angle X-ray scattering
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SOC	super optimal broth with catabolite repression
Ssu	*Streptococcus suis*
TAE	tris-acetate-ethylenediaminetetraacetic acid
TEV	tobacco etch virus
UDP	uracil diphosphate
UV	ultraviolet

Amino Acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Publications

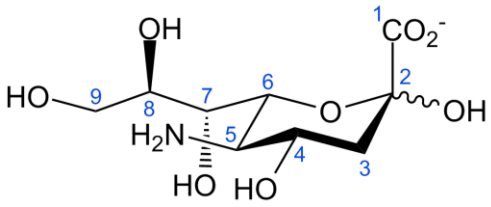
- **Joseph, D. D. A.**, Jiao, W. T., and Parker, E. J. (2013) Arg314 is essential for catalysis by *N*-acetyl neuraminic acid synthase from *Neisseria meningitidis*, *Biochemistry*, *52*, 2609-2619.
- **Joseph, D. D. A.**, Kessans S. A., Jiao, W. T., and Parker, E. J. (2014) Substrate-mediated control of the conformation of an ancillary domain delivers a competent catalytic site for N-acetylneuraminic acid synthase, *Proteins*, IN PRINT.

Chapter 1

Introduction

1.1 Sialic acids

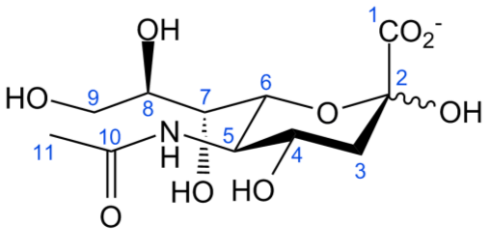
The sialic acids are a structurally diverse family of nine-carbon sugars which contain an α-keto acid functionality (*1*). Sialic acids are derivatives of neuraminic acid, which has an amino group at position C-5 and a carboxylate group at C-1 (Figure 1.1). The carboxylate group of neuraminic acid confers a negative charge under physiological conditions, making the compound a strong organic acid with a pK_a of 2.2 (*2*).



Neuraminic acid

Figure 1.1 The structure of neuraminic acid.

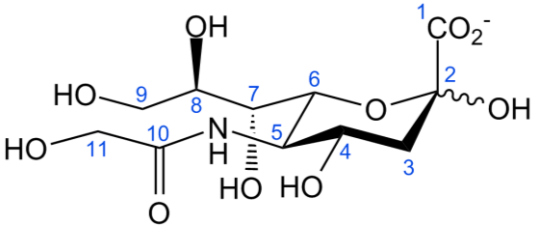
The unmodified sialic acid, neuraminic acid does not occur naturally (*2*). The most common sialic acid in nature is *N*-acetylneuraminic acid (NANA), in which the C-5 amino group of neuraminic acid is acetylated (Figure 1.2). Whilst NANA is not ubiquitous in nature it is found in almost all vestiges of life, from mammals to bacteria (*1*). NANA is so prevalent in nature, it is often synonymous with the term sialic acid (*3-6*). NANA is also a precursor for the production of many pharmaceutical drugs including Zanamivir, a neuraminidase inhibitor which has been used to treat and prevent the infection of influenza virus strains H5N1 and H1N1 (*7*).



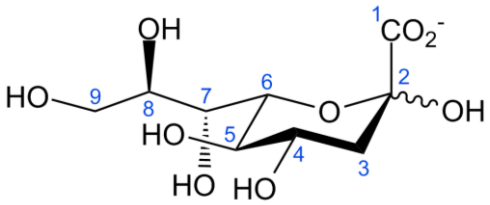
***N*-Acetylneuraminic acid**

Figure 1.2 The structure of *N*-acetylneuraminic acid.

Other common naturally occurring sialic acids include *N*-glycolylneuraminic acid (NGNA), in which the C-5 amino group of neuraminic acid is glycolylated and 2-keto-3-deoxy-nonulosonic acid (KDN), in which the C-5 amino group of neuraminic acid is replaced with a hydroxyl (Figure 1.3). NGNA is common in many animal species, however, it is not naturally present in humans due to an irreversible mutation to the gene responsible for its biosynthesis (*8*). NGNA is produced in the cytosol via NANA, where the *N*-acetyl functional group is oxidised to produce an *N*-glycolyl group (*9*). The presence of NGNA in humans has only been noted in association with particular types of cancer (*10*). KDN is not present in many members of the animal kingdom however it is common among lower vertebrates (*11, 12*). KDN is also found in bacteria and in many types of glycoconjugates such as glycoproteins, glycolipids and capsular polysaccharides (*12*).



***N*-Glycolylneuraminic acid**



2-Keto-3-deoxy-nonulosonic acid

Figure 1.3 The structures of *N*-glycolylneuraminic acid and 2-keto-3-deoxy-nonulosonic acid.

There are over fifty sialic acid molecules currently described (*1*). Whilst NANA, NGNA and KDN are the most frequently observed natural sialic acids, several alternate derivatives of neuraminic acid have been identified (*2*). Displayed in Figure 1.4 are common modifications to the neuraminic acid structure. The most frequent modification of neuraminic acid is esterification or *O*-acetylation of the hydroxyl groups, mainly at positions C-7, C-8 and C-9 although esterification at C4 is also known (*13*). Sialic acid can also contain multiple substitutions such as in *N*-7, 8, 9-tri-*O*-acetyl-*N*-acetylneuraminic acid and *N*-7, 8, 9-tri-*O*-acetyl-*N*-glycolylneuraminic acid (*2*).

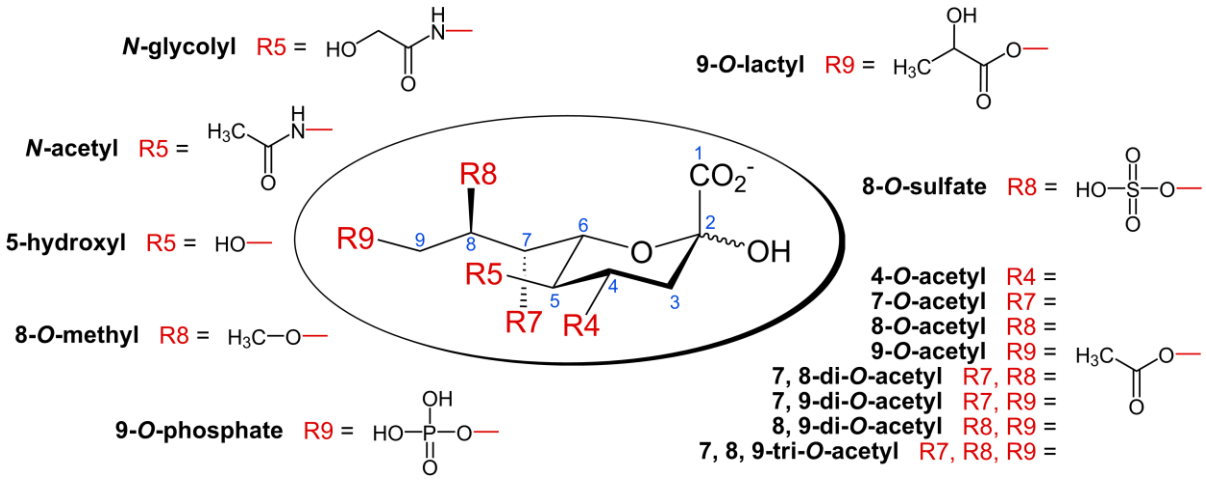


Figure 1.4 The family of naturally occurring sialic acids. Displayed in centre is the sialic acid template architecture. Figure is modified from Schauer (2004) (*1*).

1.2 Sialic acids in nature

1.2.1 Plants

The diversity of structure and functional group modifications enable the sialic acids to facilitate a vast range of cellular tasks. Whilst sialic acids are abundantly present in the animal kingdom, they have also been observed in select bacterial species and protozoa (*1*). Sialic acids have also been observed in pathogenic fungi (*14*), however, their existence in the plant kingdom is currently a topic of much debate (*15-18*).

Evidence supporting the existence of sialic acids in plant materials includes their reported quantification in several species (*18*), whilst contrary experimental analyses suggest that the levels of sialic acids present in plants are infinitesimal and within the range of extraneous contamination (*17*). The results from the latter investigation suggest that due to the small amounts of the compound measured, it is not possible to unequivocally demonstrate that plants synthesise sialic acids.

Interestingly, putative gene homologues of animal sialyltransferases and cytidine monophosphate (CMP)-NANA transporters have been detected in several plant genomes (*15, 16*). Sialyltransferases catalyse the transfer of CMP-NANA (an activated form of NANA) to the glycan moiety of glycoconjugates (*19*), whilst CMP-sialic acid transporters are typically localised in the membranes of the Golgi apparatus and provide sialyltransferases with appropriate substrate (*20*). Although a NANA metabolic pathway could not be elucidated, the putative CMP-NANA transporter-like protein in rice was observed to exhibit CMP-NANA transporter activity (*16*).

1.2.2 Fungi

The existence of NANA in fungi was first demonstrated in the yeast-like cells of *Sporothrix schenckii* (*21*) . Following its discovery in fungi, NANA has been observed in several species including *Ascocalix abietina* (*22*), *Cryptococcus neoformans* (*23*), *Fonsecaea pedrosoi* (*24*) and *Paracoccidioides brasiliensis* (*25*). Presently, NANA is the only known sialic acid observed in fungi and has only been noted in pathogenic species (*26*).

Although the metabolism and role of sialic acids in fungi are currently unknown, the incorporation of NANA in the cell surface of *A. abietina* has been shown to improve fungal resistance to unfavourable environmental conditions (*22*). Sialic acids on the cell surface of fungi are proposed to be involved with interfacial interactions (*26*). Electrostatic forces are known to mediate cell adhesion processes and surface electronegativity plays a vital role in fungal attachment to host cells (*24*). Given the strongly acidic carboxylate group of NANA (pK_a = 2.6), a NANA rich surface would provide a net negative charge under physiological conditions that would facilitate cell adhesion to positively charged substances (*26*).

1.2.3 Insects

Sialic acids were long thought to be non-existent in lower animals and insects until NANA was detected in *Drosophila melanogaster* (fruit-fly) (*27*). NANA polysaccharide was observed in the ventral nervous system and brain ganglia of *D. melanogaster* larvae. The presence of poly-NANA was only noted during a short period, 14-18 hours into development (*27*). Interestingly, NANA has not been observed in the adult form of *D. melanogaster* (*28*). Conclusive evidence for the existence of a sialic acid biosynthetic pathway in *D. melanogaster* was given upon the characterisation of an *N*-acetylneuraminic-9-phosphate synthase (NANA-9-PS) and CMP-NANA synthetase, analogous to the mammalian and bacterial versions of the enzymes respectively. (*29, 30*).

Currently, several examples indicating the presence of sialic acids in insects have been documented. Similar to *D. melanogaster*, the existence of NANA was described during the developmental stages of *Galleria mellonella* (greater wax moth) (*31*). NANA was noted to be present in the testes of *G. mellonella* with the highest quantities observed during the early pupal stages. Another example of developmental expression of NANA is in *Manduca sexta* (tobacco hornworm), where NANA was determined to be present in the inducible immune glycoprotein, scolexin (*32*). Scolexin is involved in a coagulation response to intruding bacteria (*33*). Insect glycoproteins typically terminate with either mannose or *N*-acetylglucosamine (*29*), however, sialylated glycans were observed in *Danaus plexippus* (the monarch butterfly) (*34*). NANA was also detected in glycoprotein isolated from the larvae of *Bombyx mori* (silk worm) (*35*).

1.2.4 Mammals

Whilst sialic acids are fairly scarce in lower animals, they are prolific in members of the animal kingdom from echinoderms up to higher organisms such as mammals (*36*). In mammals, NANA is incorporated as the terminal residue of cell surface glycoconjugates and plays an essential biological role in cell signalling, recognition and adhesion events (*2, 37*). Depicted in Figure 1.5 is a diagrammatic representation of a mammalian cell surface with sialylated glycoproteins.

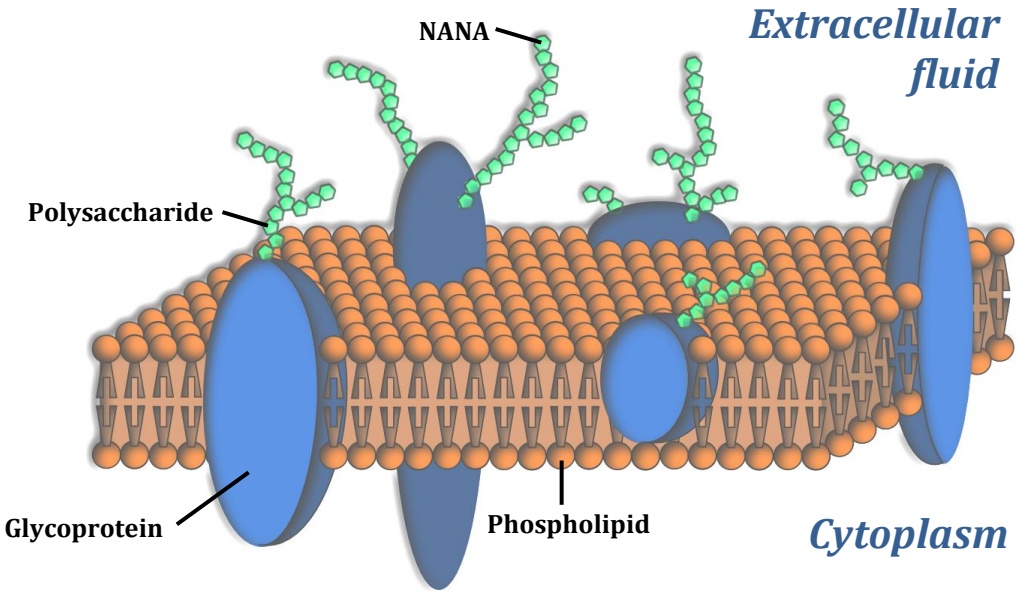


Figure 1.5 A diagrammatic representation of a mammalian cell membrane. Modified from Figure 2.6 of Medical microbiology, 4th edition (*38*)

Sialic acids are a frequent component of the oligosaccharide chains of mucins, glycoproteins and glycolipids. In mammals, the sialic acid NANA is incorporated in gangliosides (Figure 1.6) and plays a crucial role in brain development and in the normal functioning of cell membranes (*39*). Gangliosides have been shown to play an essential role in the transmission and storage of information (*40*). Increased NANA concentration in brain gangliosides has been associated with enhanced neuritic growth and has a protective effect against glutamate toxicity, where abnormally high levels of glutamate can lead to nerve cell damage and necrosis (*41*). Whilst the highest concentration of NANA is observed in brain gangliosides, it is also present in several extraneural tissues and body fluids such as saliva and milk (*40, 42, 43*).

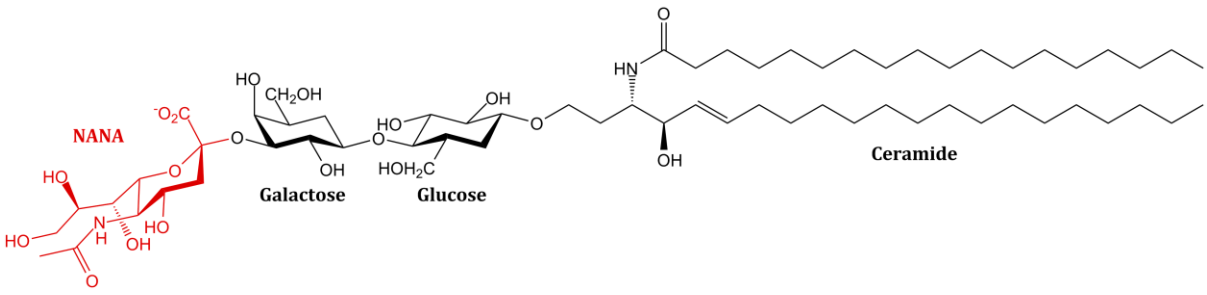


Figure 1.6 Structure of the sialoganglioside GM3 as an example of a glycoconjugate molecule (*44*). *N*-acetylneuraminic acid (NANA) is highlighted in red.

The concentration of NANA in brain gangliosides is increased in species that are higher on the phylogenetic scale (*40*). The mean concentration of NANA in the human brain was determined to be approximately 900 µg/g of ganglioside, with approximately 500 µg/g and 250 µg/g of ganglioside in rat and pig respectively (*40*).

The neural cell adhesion molecule is a heavily sialylated glycoprotein which is present in mammals and is involved in intercellular adhesion, synaptic plasticity as well as cognition and memory (*45, 46*). Gene knockout studies in mice revealed that sialylation of the neural cell adhesion molecule is essential to foetal development (*46*). The removal of enzymes that were required for the biosynthesis of NANA in mice resulted in early embryonic lethality (*46*). Furthermore, NANA is an important dietary supplement present in mammalian breast milk that is proposed to aid the rapid development of the infant brain (*45*).

As previously mentioned, NANA plays an important role in cell recognition. Sialic acid binding lectins known as siglecs interact with the specific substituents of NANA such as the carboxylate, *N*-acetyl and C-6 hydroxyl functionalities (*47*). Sialoadhesin is a particular form of siglec which is involved in the mediation of leukocyte-endothelial cell interactions in mammalian inflammatory and immune responses (*48, 49*). Sialoadhesin exists in macrophages and predominately binds neutrophils, but can also bind monocytes, natural killer cells, B cells and a subset of cytotoxic T cells by interacting with NANA molecules in the ligands on their surfaces (*49*).

NGNA is abundant in the glycoconjugates of most mammalian species, however, it is not present in healthy human tissue due to a deletion in the gene of the enzyme essential to its biosynthesis (*9, 10*). The paradoxical existence of NGNA in certain human tumour cells is attributed to NGNA absorption from dietary sources and subsequent metabolic incorporation into specific cell types (*10, 50*). Although the precise involvement of NGNA is unknown, one current hypothesis is that chronic inflammation related to the reaction between antigens and latent anti-NGNA antibodies might facilitate carcinogenesis (*50*). The net surface charge of sialic acid glycoconjugates on tumour cells produces electronegative repulsion between other cells. This results in loss of cell adhesion to the tumour mass and subsequently aids proliferation (*51, 52*).

Mucins are high molecular weight glycoproteins produced by mucosal epithelia (*43*). The peptide backbone of mucins is extensively glycosylated and a high proportion of the glycoconjugates terminate with NANA (*43*). The negative charge of NANA at physiological pH modulates bacterial and viral adhesion, however, some bacteria and viruses have evolved to colonise the mucosal niche (*43, 53*).

Neuraminidases are a class of glycosidic hydrolase enzymes which cleave the glycosidic linkages of NANA (*54, 55*), and are found in a range of organisms including mammals (*56*). The neuraminidases of certain viruses are of particular biological interest due to their role in viral pathogenicity. In particular, the viral neuraminidase of the influenza virus has been extensively researched (*54, 55, 57*). Viral neuraminidases hydrolyse the terminal NANA residues from host cell receptors and assist the mobility of virus particles via respiratory tract mucins to a target cell (*57*). Viral neuraminidases also aid in the elution of viral progeny from within an infected cell by preventing the aggregation of viral particles. Since NANA is the substrate for viral neuraminidases, several frontline drugs such as Zanamivir (Relenza™) are structural analogues of NANA (Figure 1.7) (*57*).

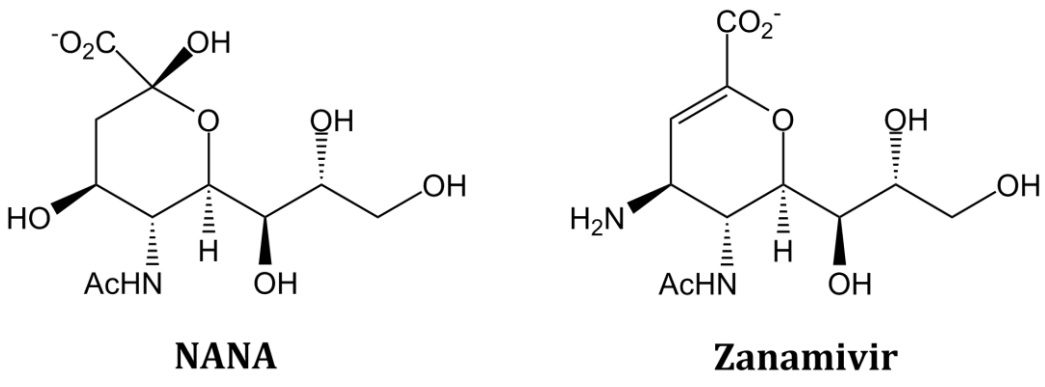


Figure 1.7 Structures of NANA and Zanamivir (Relenza™). Structural features such as the *N*-acetyl, carboxylate and glycerol functional groups are analogous to NANA (*57*).

1.2.5 Bacteria

Although sialic acids are typically absent in prokaryotes (*5*), several Gram-negative, pathogenic bacteria biosynthesise and integrate NANA into their cell surface, imitating the exterior of mammalian cells (*3*). This cell-surface mimicry enables pathogens such as *Neisseria meningitidis*, *Haemophilus influenza* and *Escherichia coli* K1, which are involved in bacterial meningitis as well as *Campylobacter jejuni*, which is involved in gastroenteritis, to evade the host immune response (*58-60*). Depicted in Figure 1.8, is a diagrammatic representation of a Gram-negative, bacterial cell wall with sialylated lipopolysaccharide.

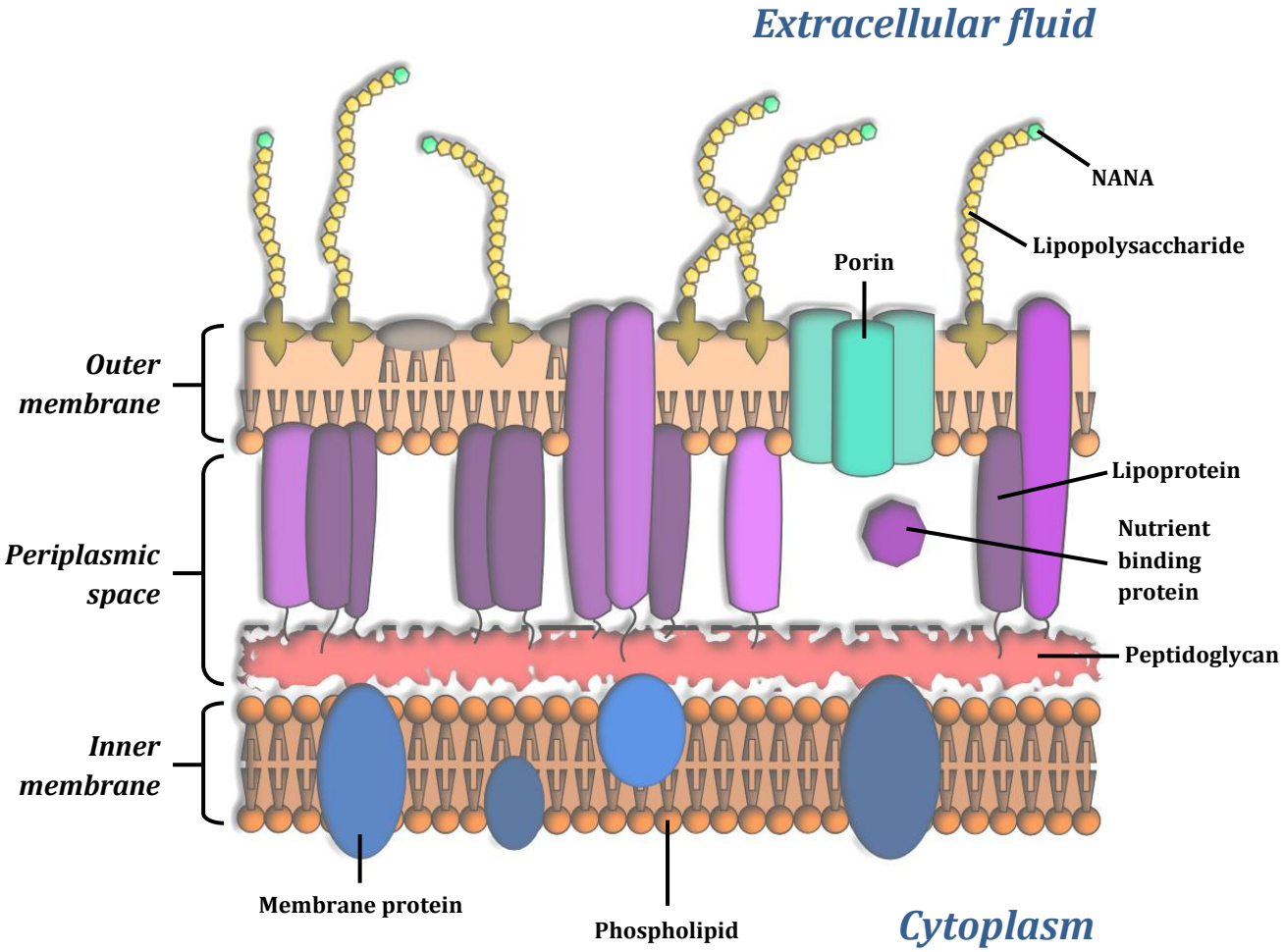


Figure 1.8 A diagrammatic representation of a Gram-negative bacterial cell wall. Modified from Figure 2.6 of Medical microbiology, 4th edition (*38*).

The sialic acid KDN is also present in several bacterial species (*12*). Bacterial KDN was first reported in the capsular polysaccharide of the Gram-negative *Klebsiella ozaenae*, a pathogenic bacterium which causes respiratory disease in humans (*61*). KDN has also been reported in the cell wall polysaccharides of *Sinorhizobium fredii*, which is involved in nitrogen fixation in certain legumes and *Pseudomonas corrugate*, the causative agent of tomato pith necrosis (*12*).

Depicted in Figure 1.9 are pseudaminic acid (Pse) and legionaminic acid (Leg), two notable nonulosonates which are commonly referred to as bacterial sialic acids in the literature (*13*). Leg has the same absolute configuration as NANA (Figure 1.2), however, Leg has a second *N*-acetyl functionality at position C-7 and does not contain a hydroxyl group at C-9. Pse is a diastereoisomer of Leg with inverted stereochemistry at positions C-5, C-7 and C-8. Pse and Leg are bacteria specific compounds and at this time have not been observed in any other form of life (*13*).

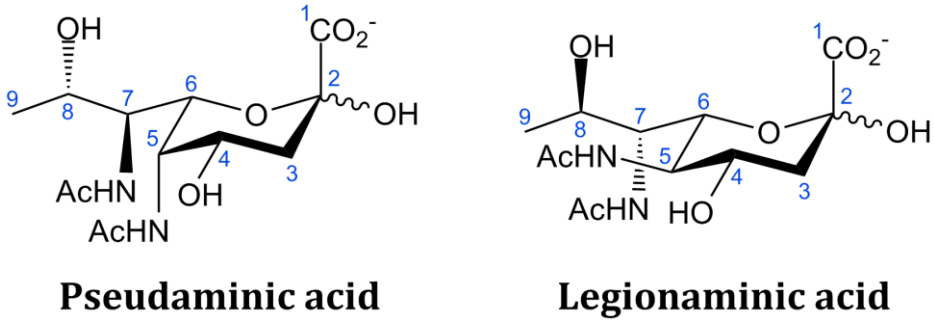


Figure 1.9 The structures of pseudaminic and legionaminic acid.

Glycosylation with Pse is an essential modification to the flagellar proteins in both *Helicobacter pylori* and *C. jejuni* (*62-64*). Flagella are complex locomotive components which are present in certain bacterial species and function by quickly rotating thin microfilaments in a reversible motion which directionally propel the microorganism (*65*). Sialylation of flagellin, the major glycoprotein component of flagella, is vital to its function (*66*). *H. pylori* and *C. jejuni* are pathogens of the gastrointestinal tract and their motility is vital to their pathogenesis (*63, 64*).

Several derivatives of Pse also occur naturally (Figure 1.10). 5-Acetamido-7-acetamidino-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid (Pse5Am5Ac) and 5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid (Pse5Ac7Am) have each been isolated from the flagellin proteins of *C. jejuni* (*67*).

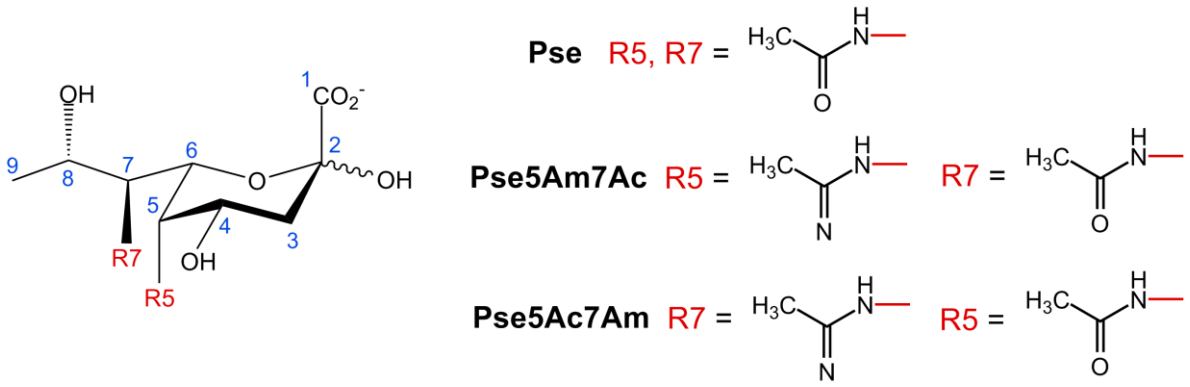


Figure 1.10 Structures of Pse and derivatives 5-acetamido-7-acetamidino-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid (Pse5Am5Ac) and 5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid (Pse5Ac7Am).

Leg is named for its presence and antigenic role in the lipopolysaccharide of *Legionella pneumophila (68)*. *L. pneumophila* is the causative pathogen of Legionnaire's disease, an often fatal pneumonia which was first observed in an outbreak at an American Legion convention in 1976 (*69, 70*). Similar to Pse, Leg has been observed as a component of flagellar glycans in both *C. jejuni* and *Campylobacter coli* (*71, 72*).

N-Methylacetimidoyl derivatives of Leg; 5-acetamidino-7-acetamido-3,5,7,9,-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid (Leg5Am7Ac) and 5-*E*/*Z*-*N*-(*N*-methylacetimidoyl)-7-acetamidino-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid (Leg5AmNMe7Ac), have also been identified in *C. coli* (Figure 1.11) (*71*). Both Leg and a novel derivative of Leg, 7-acetamido-5-(*N*-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-glycero-D-galacto-nonulosonic acid (Leg5GluNMe7Ac), have also been found in the glycosides of flagellin from *Clostridium botulinum* (*73*). *C. botulinum* is a spore-forming, Gram-positive anaerobe which produces the potent paralytic toxin botulinum neurotoxin which is responsible for food-borne botulism (*73*).

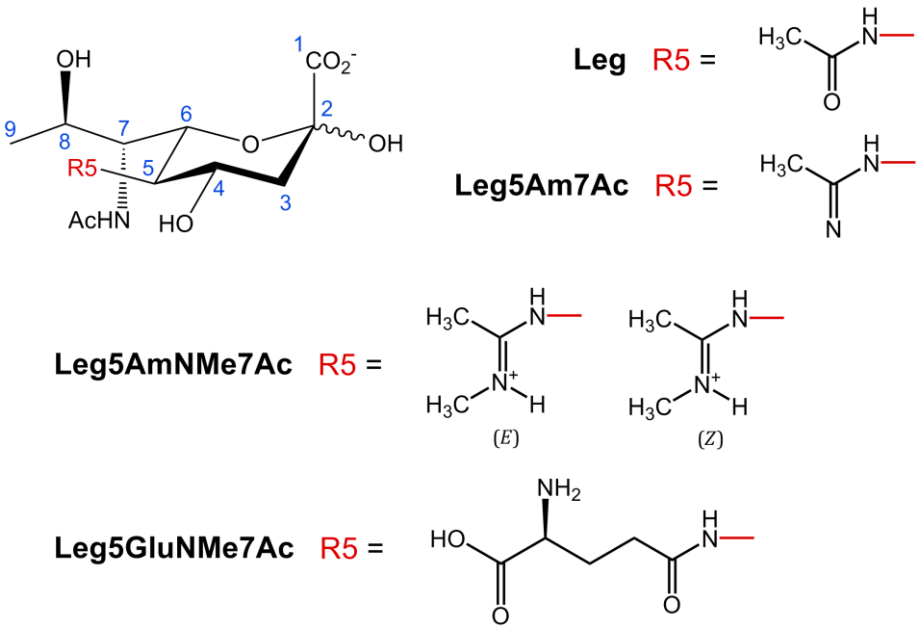


Figure 1.11 Structures of Leg and derivatives 5-acetamidino-7-acetamido-3,5,7,9,-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid (Leg5Am7Ac), 5-*E*/*Z*-*N*-(*N*-methylacetimidoyl)-7-acetamidino-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid (Leg5AmNMe7Ac) and 7-acetamido-5-(*N*-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-glycero-D-galacto-nonulosonic acid (Leg5GluNMe7Ac).

1.3 *Neisseria meningitidis*

1.3.1 Organism

N. meningitidis is a Gram-negative diplococcus bacterium and an exclusive pathogen of humans (*74*). It is frequently observed to colonise the human nasopharynx and oropharynx, however, the bacteria has also been noted to colonise the oral mucosa, rectum and urogenital tract (*74*). *N. meningitidis* has substantial genetic variation, with 13 different serogroups currently identified which are categorised by differentiation of capsular polysaccharide structures (*75*). The genetic plasticity and phenotypic diversity observed in *N. meningitidis* is speculated to be a result of recurrent interspecies recombination and horizontal gene transfer (*76*).

1.3.2 Bacterial meningitis

N. meningitidis is the leading causative agent of bacterial meningitis in the world and infection can also result in pneumonia and septicaemia (*77-79*). Meningitis is characterised by fever, headache, stiff neck, nausea, vomiting, photophobia, and altered consciousness (*80*). The annual number of deaths caused by invasive diseases in 2012 was reported to be 1.2 million, with 135, 000 deaths related to meningococcal disease (*79*). Displayed in Figure 1.12, Sub-Saharan Africa contains a region where bacterial meningitis infection is particularly rampant and has consequentially been dubbed the meningitis belt (*78*). In this particular region, the majority of meningococcal outbreaks are caused by serotype A (*78*). Conversely, in the developed world meningitis infection is primarily caused by serotype B (*74*). Meningococcal disease primarily impacts children and the fatality rate can be up to twenty times that of the adult population (*74*).

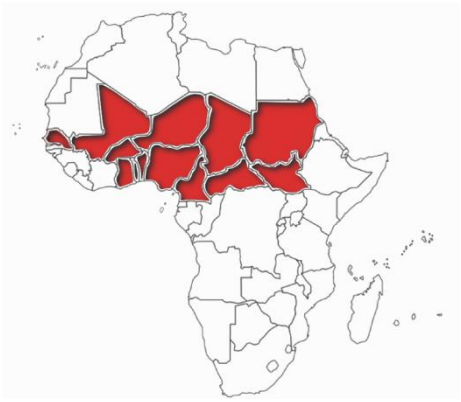


Figure 1.12 The Sub-Sahara African meningitis belt (red). This region is accountable for 65% of annual meningococcal incidences in Africa. Infections are primarily caused by serotype A followed by serotype W-135, C and X. Figure modified from Stephens *et al* (2007) (*81*).

As mentioned in the previous Section, *N. meningitidis* is most commonly observed in the human nasopharynx and oropharynx. *N. meningitidis* cells are thus transmitted via direct contact with nasal or oral secretions during activities such as coughing, sneezing or kissing (*74*). The virulence of *N. meningitidis* is associated with the components of the organism's outer membrane, specifically the capsular polysaccharide and lipo-oligosaccharide (*82*).

Colonisation by meningococci is initiated by adhesion to epithelial surfaces, which is subsequently followed by invasion of a target cell (*83*). Once internalised, the pathogen induces cytotoxicity of the host cell, which is mediated by tumour necrosis factor-α (*83*). *N. meningitidis* is capable of traversing the mucosal barrier and entering the host's blood stream which results in septicaemia. The pathogen is also capable of traversing the blood-brain barrier which is most often fatal (*83*).

The persistence of the meningitis bacterium in the host is a vital factor in its pathogenicity. *N. meninigitidis* is capable of evading the host immune response through sialylation of its lipopolysaccharide (*59*). Sialylation of the meningococcal lipopolysaccharide has been observed to inhibit phagocytosis (*82*).

As previously mentioned, there are 13 known serogroups of *N. meningitidis*, however the majority of infectious cases are resultant from the A, B, C, X, Y and W-135 serogroups (*79*). Vaccinations against serotypes A, C, W-135, and Y are currently available, whilst vaccination for serotype B is presently under development (*75*). Treatment of bacterial meningitis infections are typically through the use of antibiotics (*84, 85*). Cefalosporin-based antibiotics such as cefotaxime or cetriaxone are typically used, however, chloramphenicol (either alone or in combination with ampicillin) is known to be equally effective (*85*). The appearance of antibiotic resistance in *N. meningitidis* is of considerable concern. Meningitis strains with resistance to penicillin have been reported in Spain and resistance to chloramphenicol has also been observed in south-east Asia (*74*). Alarmingly, a recent study has shown that strain W-135 is particularly resistant to several different families of antibiotics (*86*). Given the evolutionary capacity of *N. meningitidis* and the emergence of drug-resistant strains, the development of novel therapeutics is of increased importance in the fight against meningococcal infection.

1.4 *Campylobacter jejuni*

1.4.1 Organism

C. jejuni is a motile, curved, Gram-negative bacterium (*87*). *C. jejuni* is microaerophilic and shows optimal growth in an atmosphere containing approximately 10 % CO_2 and only 5 % O_2 (*87*). The bacteria has a particularly narrow temperature growth range from a minimum of 30 °C to a maximal temperature of approximately 46 °C (*87*). *C. jejuni* produces the sialic acids NANA, Pse and Leg (*88-90*). Gene knockout studies in *C. jejuni* identified the necessity of both Pse and Leg in the natural functioning of flagella (*88*). In particular, *C. jejuni* strains without the ability to produce Pse were non-motile. Conversely, the absence of NANA biosynthesis did not affect motility (*88*).

1.4.2 Pathogenesis

C. jejuni is a food-borne pathogen of humans and is the main causative agent of gastroenteritis, although it can also cause septicaemia and meningitis as well as the neuropathological Guillain–Barré and Miller Fisher syndromes (*87, 91, 92*). *C. jejuni* infection is characterised by profuse and often bloody diarrhoea as well as acute abdominal pain and fever. Campylobacters are the principal source of bacterial diarrhoea worldwide (*91*). A reported 82% of people admitted to hospital under the diagnosis of food poisoning in the UK suffered from a campylobacter infection (*93*).

C. jejuni is primarily observed to colonise the intestinal tract of chickens, turkeys and ducks however, it has been noted in several wild and domestic animal species (*92*). The bacterium is zoonotic and able to cross between species (*91*). Human infection typically arises from consumption of uncooked food and contact with animal wastes. Many of the symptoms caused by *C. jejuni* virulence arise from inflammatory permeation of neutrophils, whilst villi atrophy and ulceration of the mucosal epithelium result from cytotoxicity of *C. jejuni* (*87*). Uncomplicated cases of *C. jejuni* infection are rarely treated with antibiotics, however, an augmented microbial resistance to certain drug types has been increasingly reported (*94*).

1.5 NANA biosynthesis

1.5.1 Bacteria

As previously mentioned, NANA is the most abundant sialic acid in nature (*1*). The biosynthetic pathway of NANA in bacteria and mammals is well documented and each of the enzymes involved have been elucidated (*5*). The route for enzymatic synthesis of NANA differs between bacterial and mammalian species (*95, 96*). Depicted in Figure 1.13 is the bacterial NANA biosynthetic pathway.

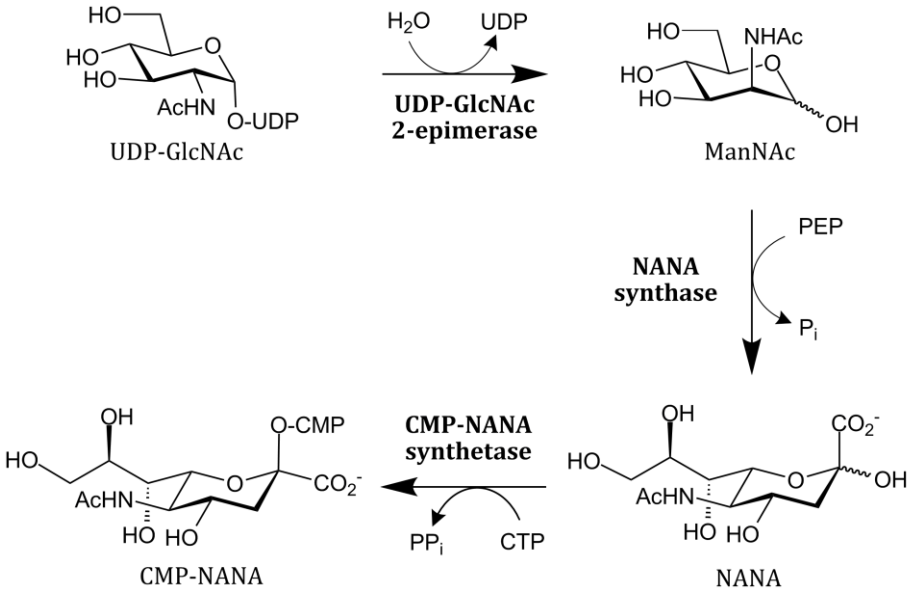


Figure 1.13 The biosynthetic pathway of CMP-NANA in bacteria. Figure modified from Tanner *et al* (2005) (*5*).

The first step in the biosynthesis of NANA is the formation of *N*-acetylmannosamine (ManNAc) from UDP linked *N*-acetylglucosamine (UDP-GlcNAc) (*97*). UDP-GlcNAc is an indirect by-product of glycolysis (*98*). The formation of ManNAc is catalysed by a hydrolysing UDP-GlcNAc 2-epimerase (*99, 100*). UDP-GlcNAc 2-epimerase catalyses an inversion at the C-2 stereocentre and hydrolyses the glycosidic phosphate bond to liberate free UDP and *N*-acetylmannosamine (*97*). The reaction catalysed by the hydrolysing epimerase is irreversible and thus the enzyme is not classified as a true epimerase (*5*).

An evolutionarily related non-hydrolysing form of UDP-GlcNAc 2-epimerase enzyme exists in bacteria, however, it is not involved in NANA biosynthesis (*101, 102*). Instead the non-hydrolysing UDP-GlcNAc 2-epimerase synthesises UDP-activated ManNAc, which is itself incorporated into the lipopolysaccharide (*102*). Unlike its hydrolysing counterpart, the non-hydrolysing UDP-GlcNAc 2-epimerase catalyses a reversible reaction (*99*).

The penultimate step in the bacterial biosynthetic pathway is an aldol-like condensation reaction between phosphoenol pyruvate (PEP) and ManNAc, catalysed by *N*-acetyl-neuraminic acid synthase (NANAS) (*103*). Bacterial enzymes with NANAS activity have been characterised from *N. meningitidis* (*104*), *C. jejuni* (*90*), *E. coli* (*95*) and *Streptococcus agalactiae* (*105*). The final bacterial reaction involves the formation of activated CMP-NANA from NANA and CTP (Figure 1.13). This reaction is catalysed by CMP-NANA synthetase (*106*). CMP-NANA is the substrate of sialyltransferase enzymes which catalyse the incorporation of NANA into cell surface glycoconjugates (*107*). Interestingly, the utilisation of CMP activated NANA is biologically unusual, as the majority of activated monosaccharides identified employ UDP or GDP linked pathways (*4*). Also of note, several bacterial species are known to utilise NANA as an alternative catabolic energy source (*59, 107*).

1.5.2 Mammals

Displayed in Figure 1.14, the CMP-NANA biosynthetic route of mammals is noticeably different to that of bacteria. Whilst the first enzymatic reaction still involves the generation of ManNAc from UDP-GlcNAc, the second step in the mammalian pathway is not the direct formation of NANA (*108*). Instead, NANA is synthesised via two phosphorylated intermediates; *N*-acetylmannosamine-6-phosphate (ManNAc-6-P) and *N*-acetylneuraminic acid-9-phosphate (NANA-9-P) (*5, 96*).

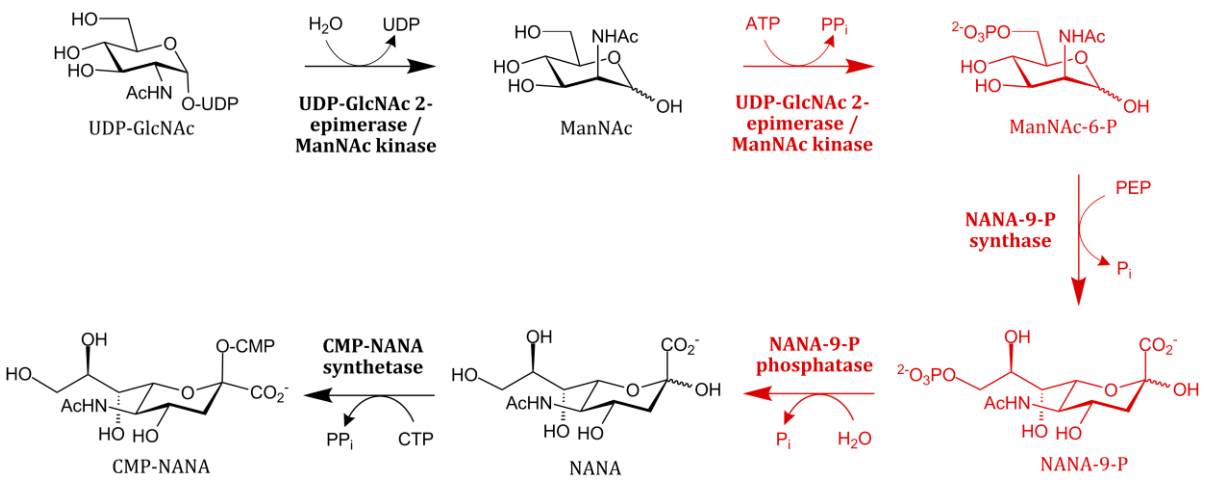


Figure 1.14 The biosynthetic pathway of CMP-NANA in mammals. Enzymatic reactions that are not present in the bacterial pathway are highlighted in red. Figure modified from Tanner *et al* (2005) (*5*).

The first and second reactions in the mammalian biosynthetic pathway are catalysed by a bifunctional enzyme with both UDP-GlcNAc 2-epimerase and ManNAc kinase activity (*109, 110*). The mammalian UDP-GlcNAc 2-epimerase/ManNAc kinase is able to convert UDP-GlcNAc to ManNAc in a similar reaction to that observed in the bacterial hydrolysing orthologue (*110*). Unlike the bacterial UDP-GlcNAc 2-epimerase, the mammalian enzyme possesses ManNAc kinase activity and is capable of condensing ATP and ManNAc to produce ManNAc-6-P and free pyrophosphate (Figure 1.14).

The third enzymatic step is analogous to that of the bacterial NANAS enzyme (*103*). Instead of condensing ManNAc and PEP, however, the mammalian orthologue condenses ManNAc-6-P and PEP to produce NANA-9-P (*103*). Depicted in Figure 1.14, NANA-9-P is subsequently dephosphorylated by NANA-9-P phosphatase to produce NANA and inorganic phosphate (*5*). Similar to the bacterial pathway, the final reaction in mammals is the formation of activated NANA via CMP-NANA synthetase (*111*).

1.6 NANA synthase

1.6.1 Enzyme structure

The structure of *Nme*NANAS is the only defined crystal structure of a NANAS enzyme currently available (*112, 113*). *Nme*NANAS has been crystallised with Mn^{2+}, PEP and a reduced form of ManNAc (rManNAc) bound in the active site (*112*), and has also been crystallised with an analogue of the reaction intermediate (*113*). *Nme*NANAS was elucidated as a domain swapped homodimer (Figure 1.15). Each monomer comprises of an N-terminal catalytic $(\beta/\alpha)_8$ barrel with an extended linker region to a C-terminal antifreeze protein-like (AFPL) domain. The N-terminal catalytic domain contributes the majority of the active site residues, whilst a functionally essential arginine residue is provided by the AFPL domain of the opposing monomer (*112, 114*).

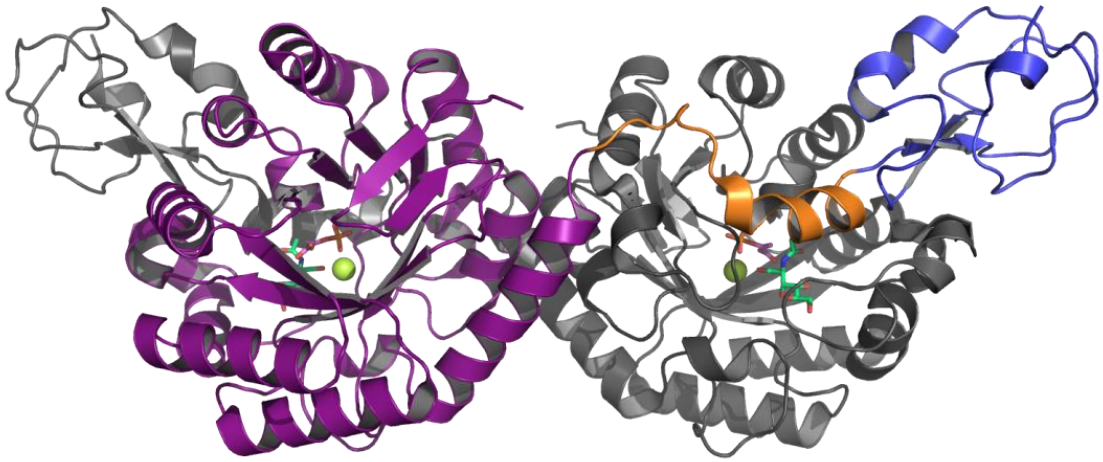


Figure 1.15 *Nme*NANAS dimer with Chain A coloured grey and Chain B coloured by domain. The catalytic domain of chain B is shown in purple (residues Q2-G271), linker domain in orange (residues G272-F290) and AFPL domain (residues A291-E349) in blue. rManNAc is coloured green, PEP pink, Mn^{2+} yellow. *Nme*NANAS PDB coordinates 1XUZ (*112*).

The AFPL domain is a unique extension which is only observed in sialic acid synthases (SASs), the family of enzymes of which NANAS belongs to (*112, 115*). Interestingly, several putative SAS enzymes lack sequence corresponding to the AFPL domain. The AFPL domain is named due to its structural homology to Type III antifreeze proteins which are commonly found in fish (*112*), and inhibit the formation of ice through thermal hysteresis (*116*). Whilst a complete structure of a NANA-9-PS enzyme is unavailable, the AFPL domain of the NANA-9-PS from *H. sapiens* has been elucidated by NMR (*115*). Displayed in Figure 1.16, the structure of the *Hsa*NANA-9-PS AFPL domain is similar to that of the *Nme*NANAS enzyme (RMSD = 0.40).

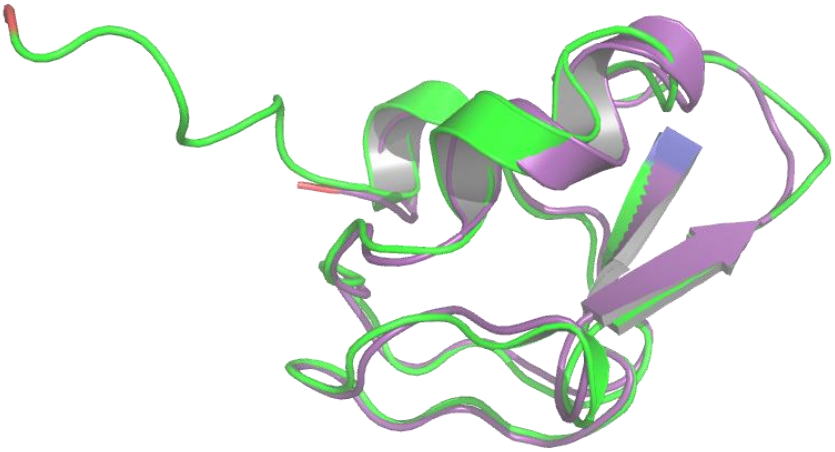


Figure 1.16 Alignment of *Nme*NANAS AFPL domain (purple) with *Hsa*NANA-9-PS AFPL domain (green). N-terminal residues are coloured blue and C-terminal residues are coloured red. *Nme*NANAS AFPL domain structure generated from crystallographic coordinates 1XUZ (*112*) and *Hsa*NANA-9-PS AFPL domain from coordinates 1WVO (*115*).

The existence of the antifreeze protein fold in bacteria and mammals has raised speculation into the genesis of this unique conformation (*115, 117*). Previous genetic analyses have suggested that the origins of both the mammalian NANA-9-PS and Type III antifreeze proteins share evolutionary roots with an ancestral SAS (*115, 117*). This is evidenced by sequence and structural analyses of several Type III antifreeze protein and AFPL domain sequences, which suggest that the Type III antifreeze proteins are more closely related to the mammalian AFPL domains than the bacterial AFPL domains (*115*). Furthermore, several of the key residues in the ice binding motif of Type III antifreeze proteins are observed in the peptide sequence of the human AFPL domain (approximately 40% sequence identity) (*115*). It has been inferred from this that the mammalian AFPL domain is an evolutionary precursor of the Type III antifreeze protein (*115*).

1.6.2 Catalytic site

Displayed in Figure 1.17 is the active site architecture of *Nme*NANAS with Mn^{2+}, PEP and rManNAc bound (*112*). As previously mentioned, the majority of the catalytic residues in *Nme*NANAS are contributed by the N-terminal domain. A single Phe288 residue is contributed by the linker region and a highly conserved Arg314 is contributed by the AFPL domain. Phe288 forms a part of the hydrophobic pocket for the ManNAc binding site, whilst Arg314 directly interacts with the *N*-acetyl carbonyl of rManNAc in *Nme*NANAS 1XUZ. Phe288 is present in select members of the bacterial SAS family whilst Arg314 is completely conserved in bacterial SAS enzymes which contain sequence pertaining to an AFPL domain. Additionally, Arg314 is not present in the sequences of mammalian NANA-9-PS AFPL domains (*115*).

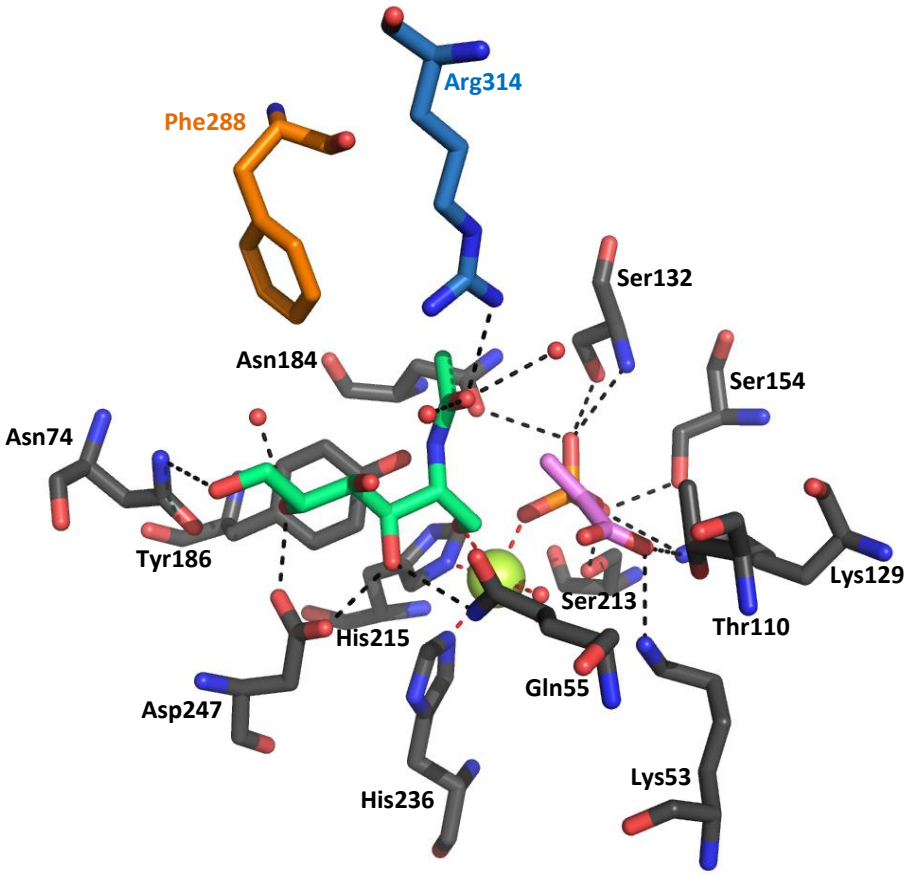


Figure 1.17 *Nme*NANAS active site. Residues contributed by the catalytic domain are in grey, linker in orange and AFPL domain in blue. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are in red. H-bonds are depicted with black dashes and dipolar bonds with red dashes (PDB code (1XUZ) (*112*)).

In addition to Arg314, the fully conserved catalytic domain residues Asp247, Gln55 and Tyr186 each interact with rManNAc (Figure 1.17). Asn74, which is contributed by the β2-α2 loop of the catalytic domain, interacts with the C-6 hydroxyl of ManNAc. Asn74 is fully conserved in NANAS enzymes but is not present in other members of the SAS family.

As depicted in Figure 1.17, Ser132, Ser154 and Ser213 form H-bonds to the phosphate group of PEP. Each of the serine residues are fully conserved in bacterial SASs and mammalian NANA-9-PSs. Asn184 also interacts with the phosphate of PEP, however, it is not completely conserved in the SAS family. The conserved residues Lys53 and Lys129 of the catalytic domain interact with the carboxylate group of PEP (Figure 1.17). The reactive group of Thr110 is proximal to the PEP carboxylate but is not within hydrogen bonding range in *Nme*NANAS 1XUZ. The positioning of Thr110 is consistent amongst available *Nme*NANAS structures (*112-114*). Additionally, Thr110 is conserved as either a threonine or a serine in the SAS family.

The active site Mn^{2+} ion has octahedral geometry and forms interactions with the imidazole side-chains of His215 and His236 (Figure 1.18). His215 and His236 are fully conserved in both mammalian and bacterial SASs, indicating that metal dependence is an essential evolutionary trait. A water molecule and the O2P of PEP complete the ligand sites of the equatorial plane whilst the axial positions of the metal ion are occupied by a second water molecule and O1 of rManNAc. The positioning of the C-1 hydroxyl group from rManNAc suggests that the carbonyl group of the non-reduced sugar substrate would be in an ideal location for electrophilic activation by the metal ion (*112*).

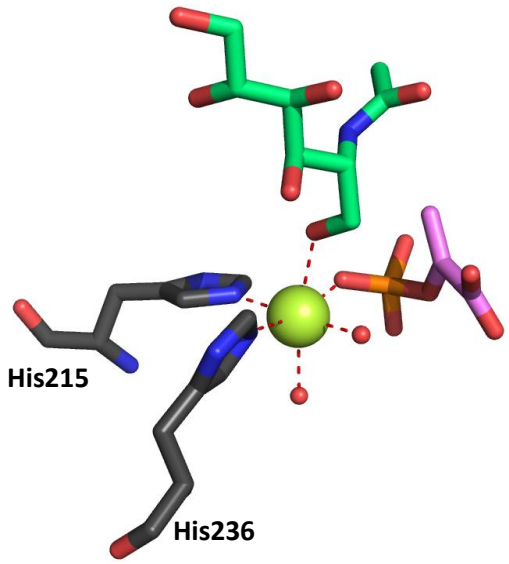
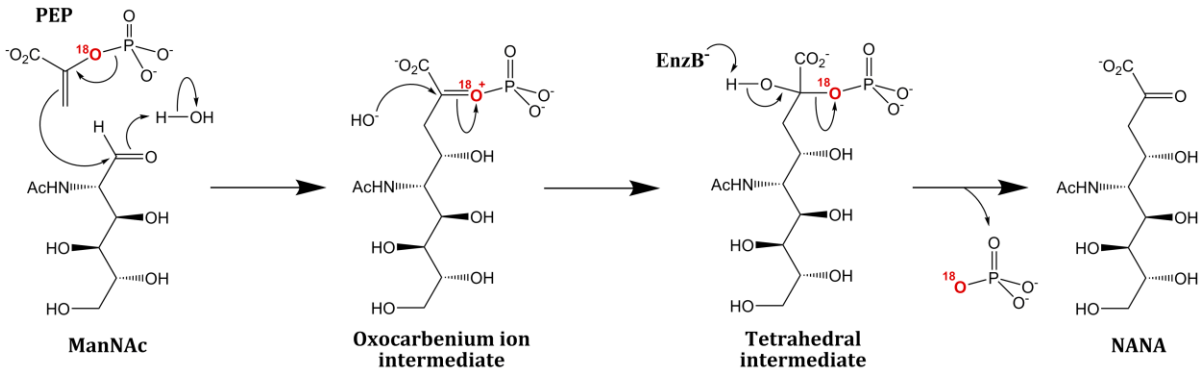


Figure 1.18 *Nme*NANAS metal coordination. Residues contributed by the enzyme are in grey. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are in red. Metal-ligand interactions are depicted with with red dashes (PDB code (1XUZ) (*112*)).

1.6.3 Catalytic mechanism

As previously mentioned, NANAS catalyses the formation of NANA via an aldol-like condensation reaction between ManNAc and PEP (*104*). NANAS is a metalloenzyme and requires a divalent metal ion for catalysis (*112*). Mn^{2+} has been established as the most activating metal ion for catalysis in both *Nme*NANAS and *Cje*NANAS followed by Co^{2+} (*90, 104*). Whilst ManNAc exists primarily in a ring-closed, pyranose form in solution, it is also in equilibrium with its open chain form. It is the open chain form that is reacted with PEP by NANAS as evidenced by ligand bound crystal structures of *Nme*NANAS (*112, 113*). Displayed in Figure 1.19 are two proposed mechanisms for the reaction catalysed by NANAS arising from nucleophilic attack by PEP.

A



B

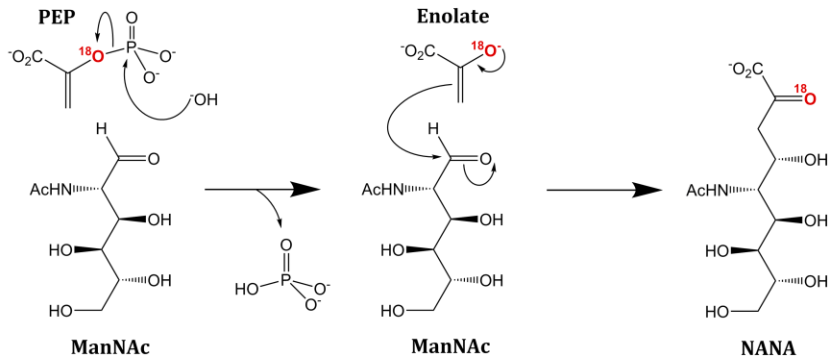


Figure 1.19 Two potential mechanisms for the reaction catalysed by NANAS enzymes. (A) A C-O bond cleavage mechanism via an oxocarbenium ion and tetrahedral intermediates. (B) A P-O bond cleavage mechanism via an enolate nucleophile. Figure modified from Tanner (2005) (*5*).

The first mechanism is a C-O bond cleavage reaction (Figure 1.19A). This is initiated by attack of the ManNAc carbonyl at C-1 by electrons from the double bond of PEP. The result of the nucleophilic attack from PEP is the formation of an oxocarbenium ion intermediate which is then readily attacked by water to yield a tetrahedral intermediate. The formation of the tetrahedral intermediate is also possible through a concerted mechanism that is not depicted. Once produced, the tetrahedral intermediate is resolved by the elimination of phosphate, affording free NANA and inorganic phosphate (*5*).

The second possible mechanism is a P-O bond cleavage reaction (Figure 1.19B). This reaction mechanism is initiated by the formation of an enolate nucleophile derived from PEP. The phosphorous group of PEP is attacked by water yielding an enolate form of pyruvate and inorganic phosphorous. The resultant enolate attacks the carbonyl of ManNAc at C-1 to produce NANA (*5*).

Mechanistic studies with isotopic labelling of PEP were used to distinguish which of the aforementioned mechanisms occurs in NANAS (*112*). An ^{18}O incorporated into the phosphate O, bridging C-2 and P in PEP was used to determine whether the C-O bond or P-O bond is cleaved during the enzyme catalysed reaction. For the mechanism where the C-O bond is cleaved, the ^{18}O label is retained in the inorganic phosphate by-product (Figure 1.19A). Conversely, in the P-O bond cleavage mechanism, the ^{18}O label is incorporated into NANA (Figure 1.19B).

The ^{18}O labelling experiment was carried out using *Nme*NANAS, and isotopic labelling of reaction products were monitored using ^{31}P NMR spectroscopy (*112*). An upfield shift of resonance resulting from a P-^{18}O bond was observed in the isolated products (*112*). This indicated that the ^{18}O label was incorporated into the inorganic phosphate and not NANA, revealing that the reaction catalysed by NANAS proceeds via a C-O bond cleavage mechanism. Furthermore, this mechanism implies that either a water or a hydroxyl group acts as a nucleophile on C-2 of PEP.

The stereospecificity of the nucleophilic attack by PEP has also been examined in previous studies with *Cje*NANAS (*90*). PEP can theoretically attack either the *si* or *re* face of the C-1 aldehyde group of ManNAc (Figure 1.19A). To determine the facial selectivity of the NANAS enzyme, the products from the reaction of ManNAc with Z and E isomers of mono-deuterated PEP were analysed using ^{1}H NMR. Depicted in Figure 1.20, the enzymatic reaction utilising *Z*-[3-^{2}H]-PEP produced (3*S*)-3-deutero-NANA with ^{2}H at the equatorial position, whereas the reaction with *E*-[3-^{2}H]-PEP produced (3*R*)-3-deutero-NANA with ^{2}H at the axial position (*90*).

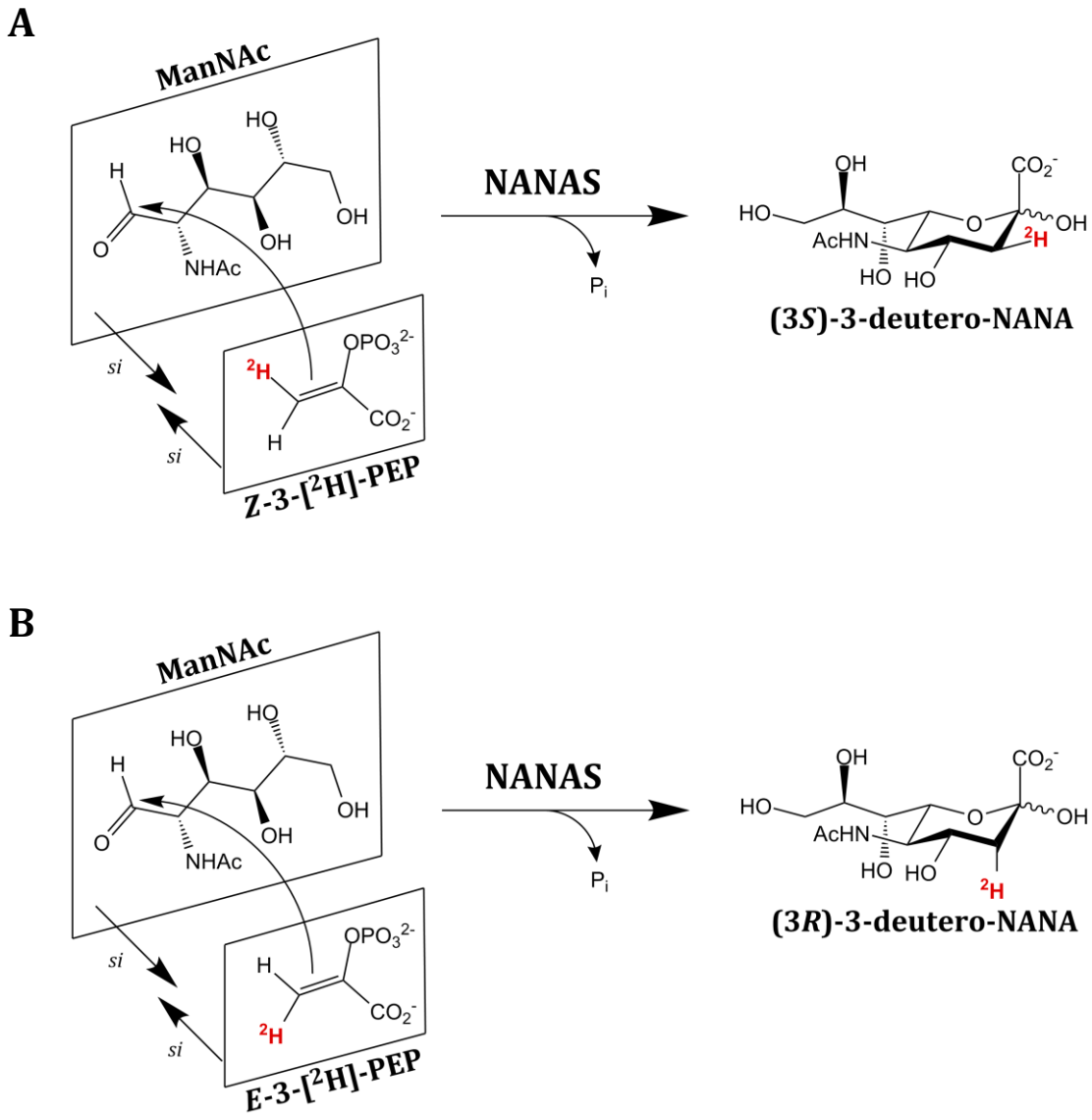


Figure 1.20 Facial selectivity of the reaction catalysed by NANAS enzymes. (A) Reaction of ManNAc with *Z*-[3-^{2}H]-PEP to produce (3*S*)-3-deutero-NANA. (B) Reaction of ManNAc with *Z*-[3-^{2}H]-PEP to produce (3*S*)-3-deutero-NANA. Figure modified from Tanner (2005) (*5*).

Similar experiments were undertaken using isomers of mono-fluorinated PEP. *Cje*NANAS was assayed with ManNAc and either *E*-[3-F]-PEP or *Z*-[3-F]-PEP, and the products were analysed using ^{19}F NMR spectroscopy (*90*). The enzyme did not accept *E*-[3-F]-PEP as a substrate whilst the reaction with *Z*-[3-F]-PEP produced (3*S*)-3-fluoro-NANA. The results from both analyses indicate that the facial selectivity of the reaction catalysed by NANAS is for the *si* face of PEP with the *si* face of the C1 carbonyl of ManNAc (*90*).

Further evidence for the *si*-facial selectivity of NANAS was afforded by the elucidation of an *Nme*NANAS crystal structure with a bound tetrahedral analogue (*113*). A structural mimic of the tetrahedral intermediate was synthesised and determined to inhibit catalysis by *Nme*NANAS (K_i of 3.1 ± 0.1 µM) (*113*). Displayed in Figure 1.21, the tetrahedral analogue bears a hydrogen instead of a hydroxyl at the C-2 stereocentre.

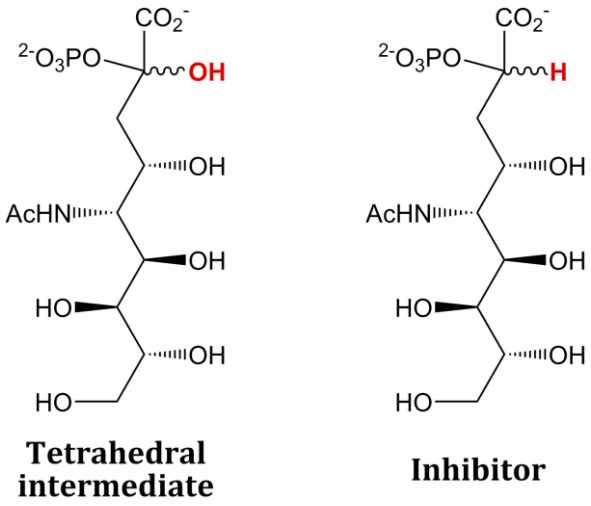


Figure 1.21 Structures of the tetrahedral intermediate for the NANAS reaction and the inhibitor based on its structure.

NANAS activity was analysed using two diastereoisomers of the intermediate mimic, with either (2*R*) or (2*S*) configuration at the C-2 tetrahedral centre. Whilst the two diastereoisomers could not be completely isolated, it was successfully demonstrated that the (2*R*) configuration accounted for the majority of the inhibition observed (*113*). Furthermore, a 1.75 Å resolution crystal structure of *Nme*NANAS was elucidated with (2*R*)-inhibitor bound in the active site even though the crystals were soaked in a 4.5:1 mixture of (2*S*)-inhibitor:(2*R*)-inhibitor (Figure 1.22) (*113*).

The preference for the (2*R*)-inhibitor observed in the *Nme*NANAS 2WQP crystal structure and improved inhibition values compared to the (2*S*)-inhibitor, likely indicates that the natural tetrahedral intermediate would share (2*R*) geometry. The (2*R*)-configuration indicates that the reaction mechanism occurs via a *si* face attack to the oxocarbenium ion intermediate. The structure may also imply that the metal ion plays a dual catalytic role; activating the aldehyde of ManNAc and acting as a source of nucleophilic, activated water for attack of the oxocarbenium ion intermediate (Figure 1.19A) (*113*).

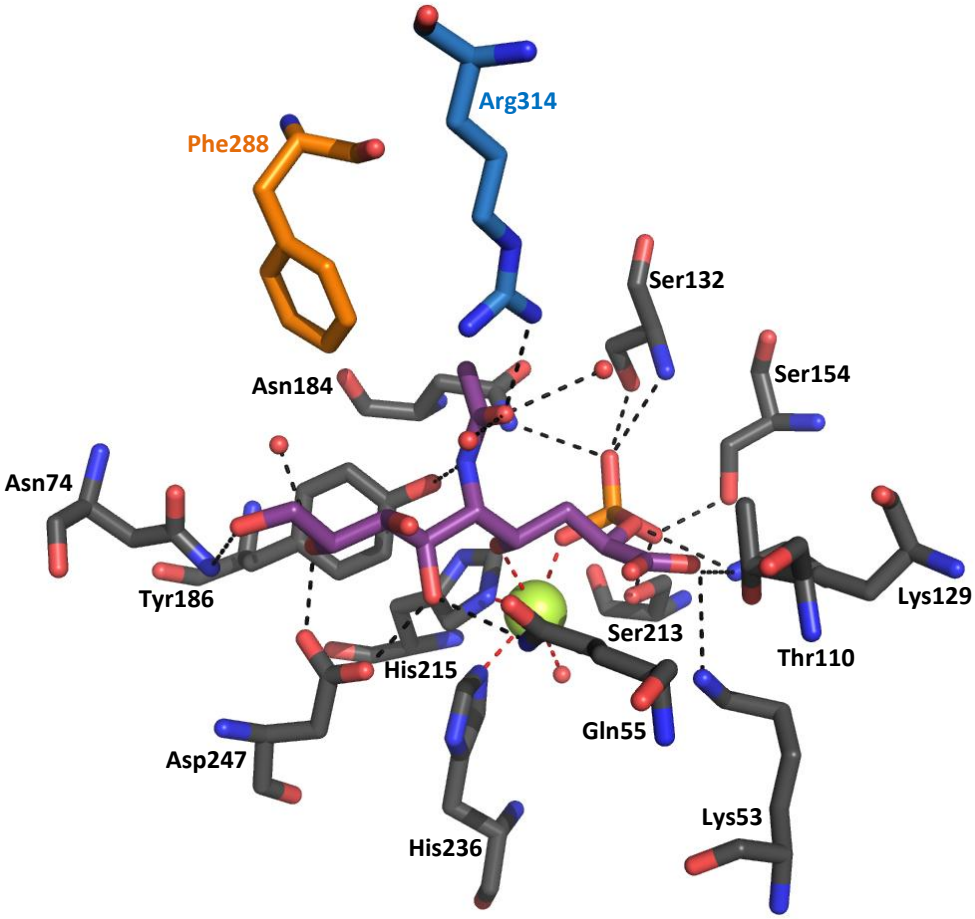


Figure 1.22 *Nme*NANAS active site with tetrahedral intermediate mimic bound. Residues contributed by the catalytic domain are in grey, linker in orange and AFPL domain in blue. Tetrahedral intermediate mimic is in purple, Mn^{2+} yellow and waters are in red. Polar contacts are depicted with black dashes and dipolar bonds with red dashes (PDB code (2WQP) (*113*)).

1.7 Evolutionary relationships

1.7.1 PEP aldolases

The C-O bond cleavage mechanism employed by the NANAS enzyme is unusual and very few PEP utilising enzymes are known to react in this manner (*5*). Enzymes which have similar catalytic mechanisms include 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA), 2-keto-3-deoxy-D-*arabino*-heptulosonate synthase (DAH7PS) and 2-keto-3-deoxy-D-*manno*-octulosonate synthase (KDO8PS) (*118-121*). The aforementioned enzymes are divided into two categories, enolpyruvyl transferases (EPSPS and MurA) and PEP aldolases (DAH7PS, KDO8PS and NANAS) (*122*). Enolpyruvyl transferases catalyse an addition-elimination reaction where a pyruvyl moiety from PEP is transferred to an acceptor alcohol (*120, 121*). In comparison, PEP aldolases catalyse an aldol-like addition of PEP to an aldehyde (*118, 119*).

The catalytic mechanism of both DAH7PS and KDO8PS are extensively documented in the literature (*118, 123, 124*). DAH7PS and KDO8PS each catalyse aldol-like reactions combining PEP with D-erythrose 4-phosphate (E4P) or D-arabinose 5-phosphate (A5P) respectively, to produce the corresponding sugar phosphate products (Figure 1.23). Due to the similarity in mechanisms employed, it has been previously proposed that the contemporary families of PEP aldolases evolved from a common ancestral enzyme (*122*). Although DAH7PS and KDO8PS have low sequence similarity to NANAS enzymes (< 10% identity) (*113*), each share a similar $(\beta/\alpha)_8$ barrel catalytic domain fold (*124, 125*). A notable difference is that in both DAH7PS and KDO8PS enzymes, the PEP molecule attacks the *re* face of the aldehyde. As previously discussed in Section 1.6.3, in NANAS it is the *si* face of the aldehyde that is attacked (*5*).

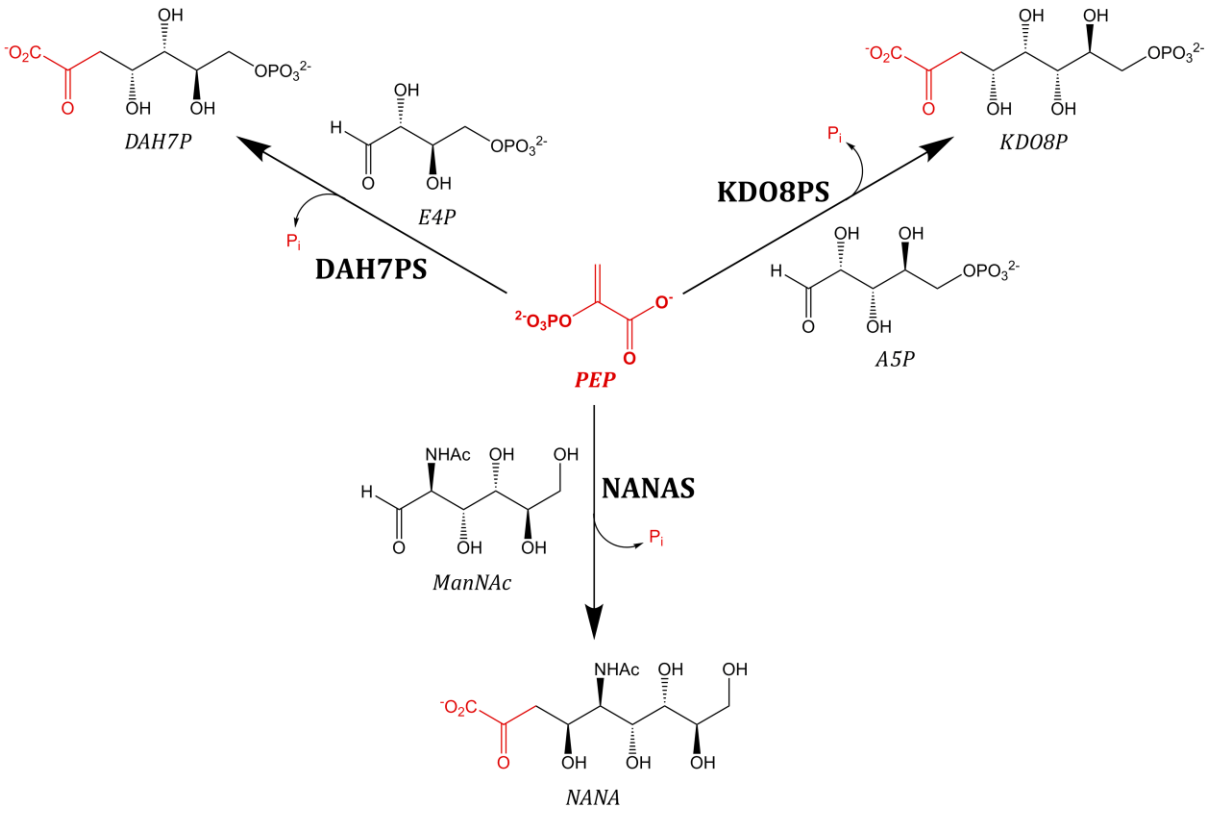


Figure 1.23 Reactions catalysed by the PEP aldolase super-family of enzymes. The reactions of DAH7PS, KDO8PS and NANAS enzymes are depicted with PEP highlighted in red.

1.7.2 DAH7PS

DAH7PS catalyses an aldol-like condensation between PEP and E4P to produce DAH7P. DAH7PS catalyses the first reaction of the shikimate pathway, the metabolic route by which chorismate is formed (Figure 1.24) (*126*). Chorismate is an important metabolite from which the aromatic amino acids, folate and vitamin K are generated (*127*). The shikimate pathway is present in plants and microorganisms but is absent in higher life forms (*126*). Enzymes from this pathway are therefore potential targets for the development of antimicrobials and herbicides. Since the reaction catalysed by DAH7PS is the first committed step in a branched enzymatic pathway, it is a key metabolic control point and is highly regulated (*127*).

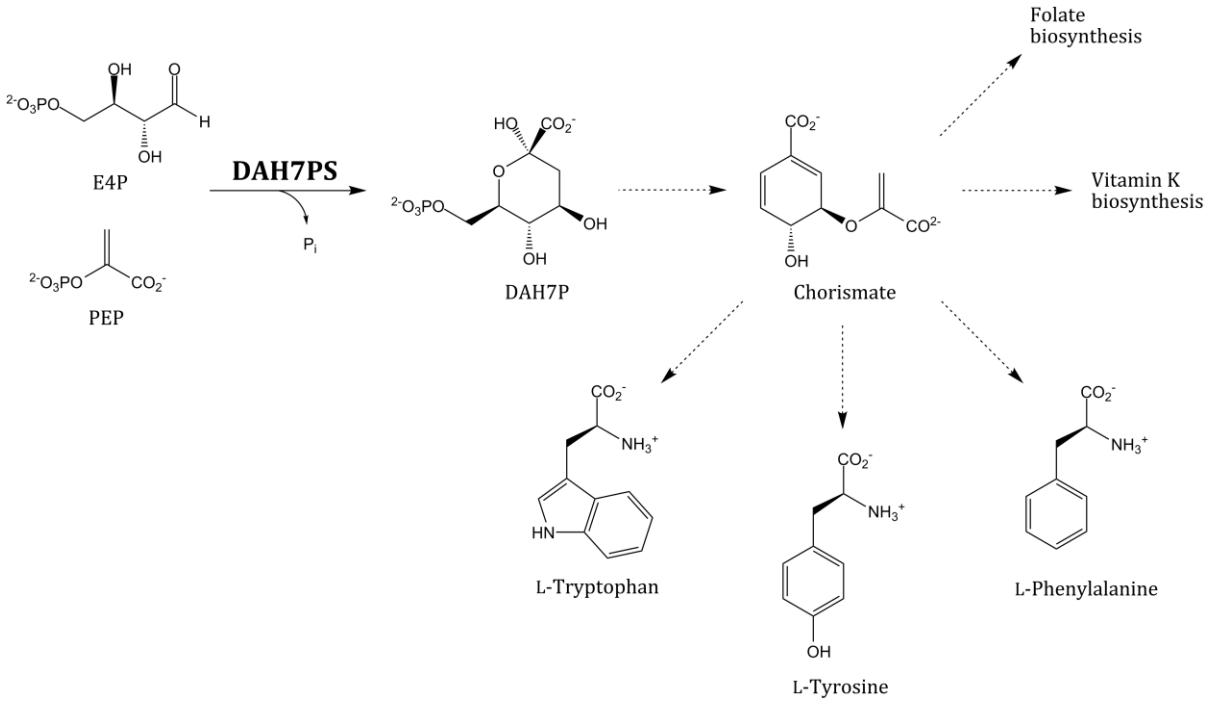


Figure 1.24 The shikimate pathway. The first reaction is catalysed by DAH7PS. Multiple enzymatic steps convert DAH7P to chorismate, which is the branch point of the aromatic amino acid, folate and vitamin K biosynthetic routes.

DAH7PS enzymes contain a core, catalytic $(\beta/\alpha)_8$ barrel yet are structurally diverse due to a variety of extra-domain decorations. These barrel extensions enable feedback regulation of the enzyme (*127*). DAH7PSs are classed as Iα, Iβ or II based on the nature of the additions (*125, 128*). Similar to NANAS enzymes, DAH7PS require a divalent metal ion for catalysis (*129*). As previously mentioned, NANAS has a low sequence identity to DAH7PS enzymes. Correlatively, the metal binding scaffolds of NANAS and DAH7PS are disparate.

1.7.3 KDO8PS

KDO8PS condenses PEP and A5P to form KDO8P (Figure 1.23), which is a metabolic precursor of 3-deoxy-D-*manno*-octulosonate (KDO); an essential component of the lipopolysaccharides of Gram-negative bacteria (*130*). Gene knock-out studies have indicated that KDO8PS performs an essential function for bacterial homeostasis, and that loss of KDO8PS function can arrest cell growth (*131, 132*). KDO8PS is thus a potential drug target against Gram-negative, pathogenic bacteria (*128*). KDO8PS enzymes are closely related to DAH7PSs, however, unlike DAH7PS and NANAS they have variable metal dependence (*113*). Several KDO8PS enzymes do not require metal ions for their catalytic function (*133*). Additionally, whilst KDO8PS, DAH7PS and NANAS share a common $(\beta/\alpha)_8$ barrel scaffold, KDO8PS enzymes do not possess extra barrel decorations (*130*).

1.7.4 Sialic acid synthase family

Whilst it was originally believed that the NANAS enzyme was evolutionarily similar to DAH7PS and KDO8PS (*122*), more recent investigations have indicated that this relationship is quite distant (*113, 134*). NANAS belongs to a family of SAS enzymes which include pseudaminic acid synthase (PseS) and legionaminic acid synthase (LegS) (*89, 135*). Displayed in Figure 1.25, LegS catalyses an aldol-like condensation reaction between PEP and 2,4-diacetamido-2,4,6-trideoxymannose to form Leg, whilst PseS reacts PEP with 2,4-diacetamido-2,4,6-trideoxyaltrose to form Pse. Genes encoding NANAS, LegS and PseS are present in the *C. jejuni* genome, making it an excellent model for studying the relationships between these three homologous enzymes (*64*).

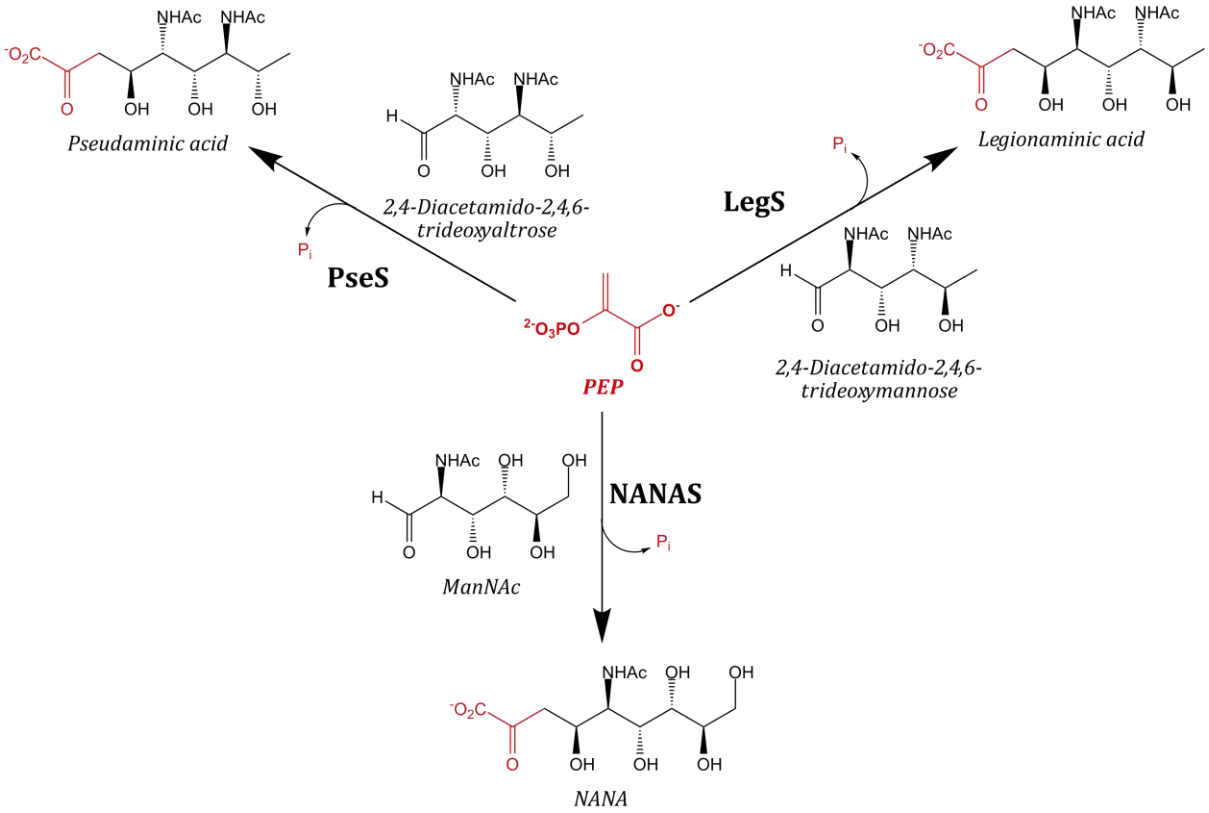


Figure 1.25 Reactions catalysed by the SAS family of enzymes. The reactions of NANAS, PseS and LegS and enzymes are depicted with PEP highlighted in red.

1.7.5 CMP-Leg biosynthesis

The biosynthetic pathways of Leg and Pse are very similar to that described for NANA in Section 1.5.1. A 30 kb locus has been previously identified in *L. pneumophila* that contains many of the necessary genes for Leg biosynthesis (*136*). Within this locus there are homologues of each of the final three enzymes involved in CMP-NANA biosynthesis (*68, 135*). As shown in Figure 1.26, the first of these enzymes is a nucleoside diphosphate (NDP)-sugar hydrolase/2-epimerase that catalyses both an inversion of configuration at C-2 and a hydrolysis of the glycosidic-NDP linkage, converting NDP-2,4-diacetamido-2,4,6-trideoxy-glucopyranose into 2,4-diacetamido-2,4,6-trideoxymannose and free NDP. The second enzyme is LegS which, as previously mentioned, condenses PEP and 2,4-diacetamido-2,4,6-trideoxymannose to form Leg and free, inorganic phosphate. The final enzyme in the CMP-Leg biosynthetic pathway is CMP-Leg synthetase which activates Leg as a CMP derivative.

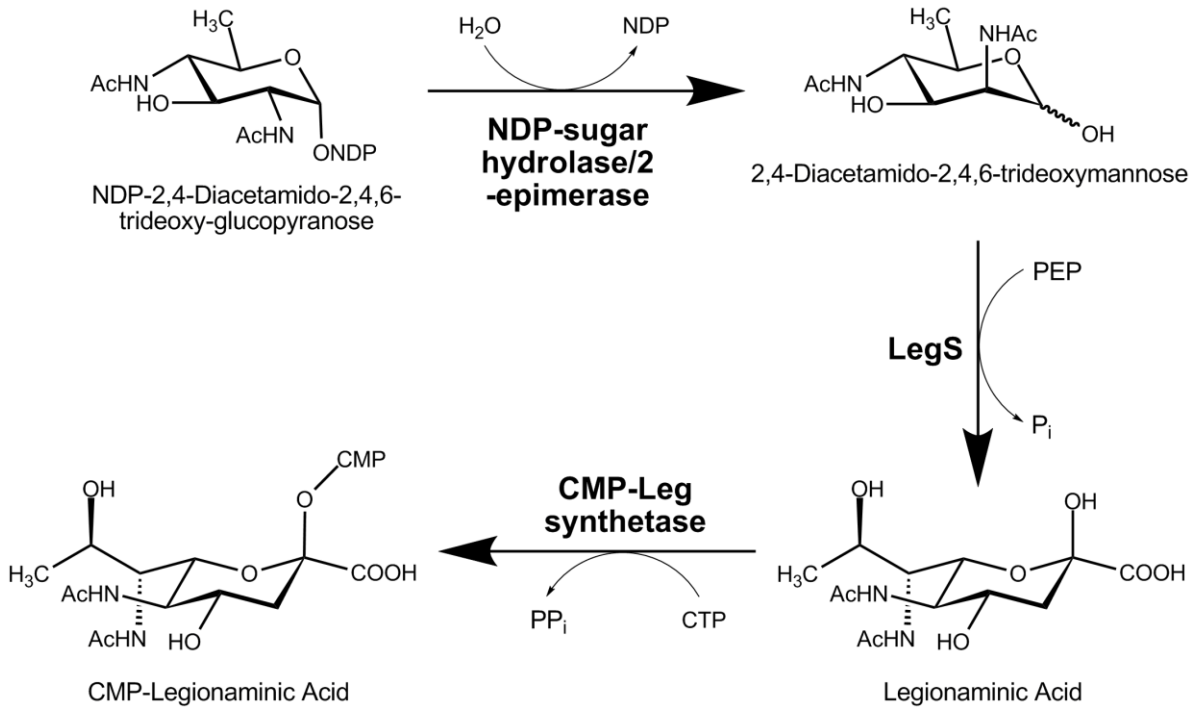


Figure 1.26 The final three enzymatic reactions of the CMP-Leg biosynthetic pathway. The substituent NDP can be either uracil (UDP) or guanine (GDP) (*68, 135*).

LegS enzymes have been previously characterised from both *L. pneumophila* (*135*) and *C. jejuni (68)*. As reported by Glaze *et al* (2008), *Lpn*LegS was characterised as a maltose binding (MalE) fusion protein as there was significant difficulty in producing soluble enzyme (*68*). *Lpn*LegS-MalE was found to catalyse the condensation reaction between PEP and 2,4-diacetamido-2,4,6-trideoxymannose through NMR analysis but had a very poor specific activity of 4.4×10^{-3} µmol min^{-1}. It was postulated that the low specific activity ascertained for *Lpn*LegS was not an accurate representation of the *in vivo* parameter due to the MalE modification. Furthermore, it was revealed by [2-^{18}O]PEP that LegS catalysed its reaction through C-O bond cleavage similar to the mechanism previously shown for NANAS (*112*).

The CMP-Leg pathway from *C. jejuni* was elucidated by Shoenhofen *et al* (2009) (*135*). It was proven by NMR that *Cje*LegS was able to condense PEP with 2,4-diacetamido-2,4,6-trideoxymannose to form Leg. Shoenhofen *et al* (2009) proposed that the biosynthetic route for CMP-Leg synthesis occurred via GDP-linked sugars as opposed to UDP-linked sugars, contrary to what had been previously reported for both NANA and Pse biosynthesis (*5, 63*). It was reasoned that the use of GDP-linked sugars enabled differentiation of the Leg biosynthetic pathway from other UDP-linked pathways such as those involving UDP-2,4-diacetamido-2,4,6-trideoxy-glucopyranose (involved in *N*-linked protein glycosylation). In comparison, UDP-linked sugars were used in the previously mentioned study by Glaze *et al* (2008), where UDP-2,4-diacetamido-2,4,6-trideoxyglucopyranose was shown to be utilised by the Leg biosynthetic sugar hydrolase/2-epimerase (*68*). It is unclear from these reports whether UDP or GDP linked precursors are preferred by the enzymes of the CMP-Leg biosynthetic pathway and it is likely that either may be utilised.

1.7.6 CMP-Pseudaminic acid biosynthesis

The CMP-Pse pathway has been elucidated in both *H. pylori* (*63*) and *C. jejuni (137, 138)*. The final three enzymes involved in CMP-Pse biosynthesis are shown in Figure 1.27. The first enzyme shown is a UDP-sugar hydrolase that catalyses a hydrolysis of the glycosidic-UDP linkage, converting UDP-2,4-diacetamido-2,4,6-trideoxyaltrose into 2,4-diacetamido-2,4,6-trideoxyaltrose. This reaction differs from the analogous step in Leg and NANA biosynthesis, where the corresponding sugar hydrolases have an additional epimerase functionality that inverts the stereochemistry at position C-2. In Pse biosynthesis, C-2 stereochemistry is retained during this step. The crystal structure of the pseudaminic UDP-sugar hydrolase has been elucidated from *C. jejuni (137)*.

The second enzyme depicted is PseS which condenses PEP and 2,4-diacetamido-2,4,6-trideoxyaltrose to form Pse and free, inorganic phosphate. *Cje*PseS has been previously characterised using a continuous coupled phosphate assay with purine nucleotide phosphorylase and product formation was confirmed by both ^{1}H and ^{31}P NMR (*89*). Using the coupled assay, *Cje*PseS was determined to have a k_{cat} = 0.65 ± 0.01 s^{-1}, K_m (PEP) = 6.5 µM and K_m (2,4-diacetamido-2,4,6-trideoxyaltrose) = 9.5 ± 0.7 µM (*89*). The most activating metal for *Cje*PseS was determined to be Co^{2+} followed by Mn^{2+}. Similar to NANAS and LegS, [2-^{18}O] PEP was used to confirm that PseS catalysed Pse synthesis via C-O bond cleavage. The final enzymatic step in the CMP-Pse biosynthetic pathway is catalysed by a CMP-synthetase which activates Pse.

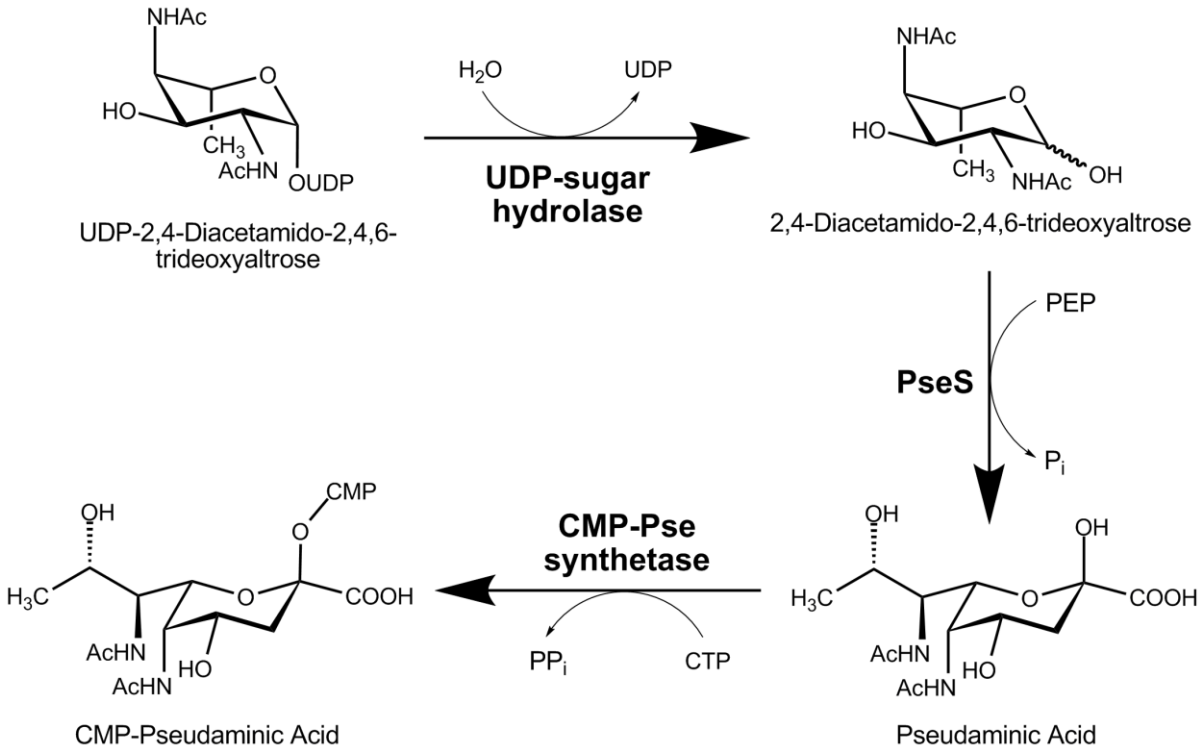


Figure 1.27 The final three enzymatic reactions of the CMP-pseudaminic acid biosynthetic pathway.

1.8 Research objectives

Sialic acids play important roles in many biological processes such as cell signalling, recognition and adhesion (*2, 37*). They are also involved in the pathogenesis of several invasive species (*59*), thus the enzymes involved in their synthesis are of substantial therapeutic interest. SASs are critical in the generation of sialic acids, and understanding their catalytic function is of significant importance to the development of novel antimicrobials. Current knowledge of SAS catalysis is limited to the reaction mechanism, and minimal investigation has been made into the role of specific catalytic site residues. Furthermore, the role of the unique AFPL domain has not been examined beyond its contribution of residue Arg314.

The evolutionary relationships within the SAS family are also poorly understood, with several SAS enzymes misannotated in online databases. In general, this thesis aims to further develop the current understanding of SAS evolutionary relationships and improve existing knowledge of the catalytic mechanism by investigation of key residues as well as the unique SAS AFPL domain.

The specific research objectives for this thesis are as follows:

- Characterisation of wild-type *Nme*NANAS using an uncoupled, continuous assay.
- Investigation of the catalytic function of *Nme*NANAS Arg314 through site-directed mutagenesis.
- To develop an understanding of the role of the AFPL domain in catalysis through the generation of an AFPL domain truncation variant.
- To use bioinformatic analysis to explore the evolutionary relationships within the SAS family and to identify key residues and sequences that may confer substrate selectivity.
- Characterisation and comparison of NANAS, LegS and PseS from *C. jejuni* to gain further insight into the relationships between these enzymes.
- Characterisation of the putative SAS enzyme from *Chlorobium tepidum* which does not natively express an AFPL domain, and developing an understanding of its evolutionary relationship to other members of the SAS family.
- Determining substrate selectivity of the SASs using an assortment of ManNAc analogues.

- Characterisation of the putative SAS enzyme from *Streptococcus suis* in an effort to observe substrate plasticity.
- Elucidation of additional SAS enzyme structures, to further study the evolutionary variation of the active site architecture.
- Generation of two AFPL domain chimeras combining the NANAS enzyme from *N. meningitidis* and NANA-9-PS from *Homo sapiens* in an effort to interchange substrate selectivity.

Since *Nme*NANAS is the most thoroughly investigated and the only SAS from which a complete structure is elucidated, it was used as a basis and point of comparison for much of the work outlined in this thesis (*5, 103, 104, 112, 113*). The SAS enzymes from *C. jejuni* were used to compare evolutionary relationships between NANAS, LegS and PseS, as a *C. jejuni* paralogue from each clade has been previously identified (*64*). The putative SAS enzyme from *C. tepidum* was selected to investigate a naturally truncated SAS. Substrate plasticity has been observed in *Streptococcus agalactiae* NANAS (*105*) and thus to compare this phenomenon, the SAS from *S. suis* (which has the highest sequence identity to *Sag*NANAS) was isolated.

Chapter 2

Investigating the role of Arg314 in *Nme*NANAS

2.1 Background

2.1.1 Structure of *N*-acetylneuraminic acid synthase

The structure of *N. meningitidis* NANAS (*Nme*NANAS), has been previously elucidated in the presence of Mn^{2+}, PEP and a reduced form of ManNAc (Figure 2.1), and has also been crystallised in the presence of a tetrahedral intermediate mimic (*112, 113*). Both structures have revealed that *Nme*NANAS is a domain swapped homodimer, with each subunit comprising of two structurally distinct domains joined by a linker region. The N-terminal domain of *Nme*NANAS is a $(\beta/\alpha)_8$ barrel fold and contributes the majority of the residues comprising the catalytic site. The C-terminal domain comprises 58 amino acid residues and strongly resembles the fold of fish Type III antifreeze proteins (*103, 117*).

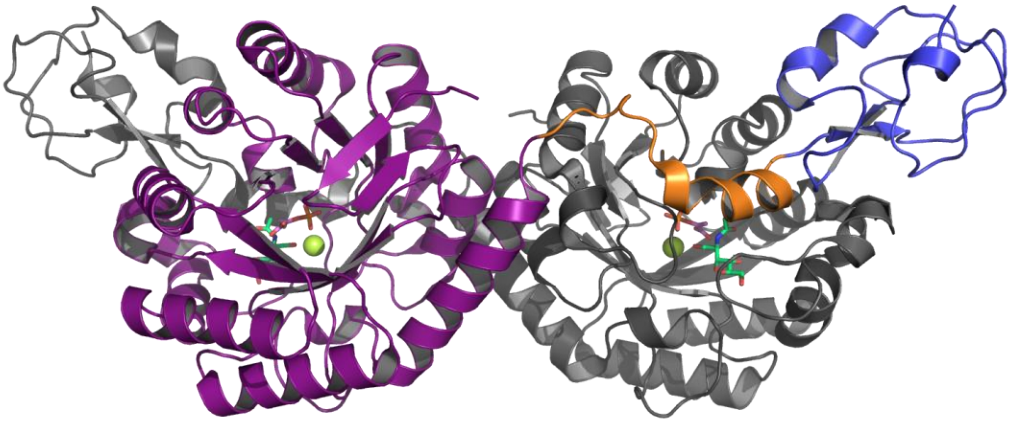


Figure 2.1 (A) *Nme*NANAS dimer with one monomer coloured grey and the second coloured by domain. The catalytic domain is shown in purple, linker domain in orange and anti-freeze protein like (AFPL) domain in blue. Crystallographic coordinates 1XUZ (*112*).

The antifreeze protein like (AFPL) domain caps the catalytic domain of the opposing monomer, donating one key residue (Arg314), directly into the active site (Figure 2.2). Arg314 is positioned to form a hydrogen bond with the acetyl group oxygen of rManNAc and is thus proposed to interact with the congruent functionality of the natural substrate. Residue Arg314 is also highly conserved in bacterial SASs. The importance of Arg314 has been previously implicated in *Streptococcus agalactiae* NANAS where treatment of the enzyme with the Arg modifying reagent phenylglyoxal was found to inhibit catalytic activity (*105*).

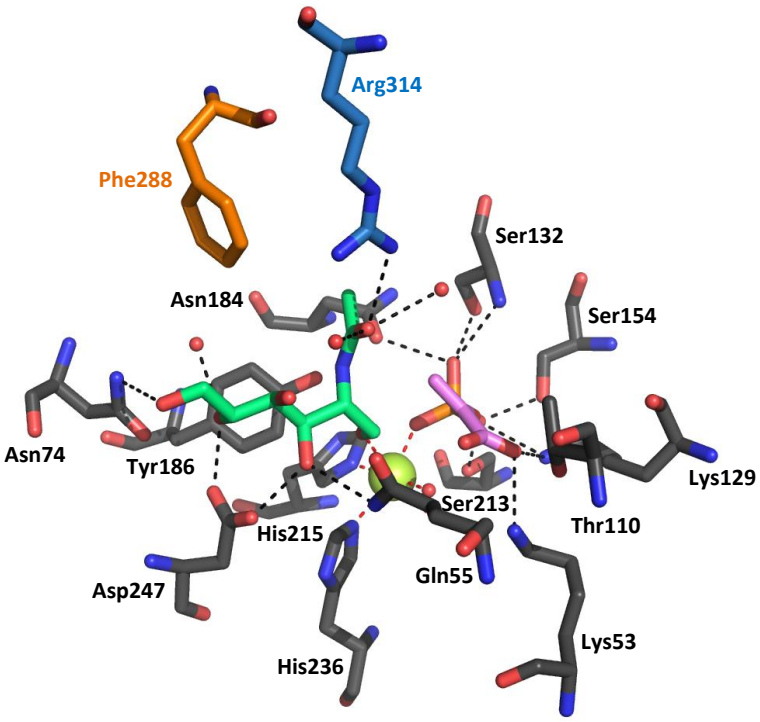


Figure 2.2 Close up of active site. Residues contributed by the catalytic domain are shown in grey and residues from linker in orange and AFPL domain in blue. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are shown in red. H-bonds are depicted with black dashes and dipolar bonds with red dashes (PDB code (1XUZ) (*112*)).

2.1.2 Goals of this research

*Nme*NANAS is the most thoroughly investigated member of the SAS family (*5, 104, 112, 113*). *Nme*NANAS is also the only wild-type SAS with a complete crystal structure available (*112, 113*). The current knowledge on this enzyme makes *Nme*NANAS an excellent starting point and model for mechanistic studies and evolutionary comparisons to other SASs. The work done in this chapter aims to build upon previous studies and to develop a point of reference for comparisons made in later chapters.

Currently very little is known about the role of the AFPL domain other than its contribution of Arg314. The work outlined in this chapter aims to investigate the role of Arg314 in the catalytic mechanism of *Nme*NANAS.

2.2 Cloning, expression and purification of *Nme*NANAS

2.2.1 Cloning

The gene encoding NANAS from *N. meningitidis* serotype B strain MC58 (*neuB*) was amplified from genomic DNA using the nested PCR method outlined in Section 8.3.2. Gene specific nested primers were designed incorporating generic extensions for further rounds of nested PCR (Table 8.5). Following successful amplification as determined by agarose gel electrophoresis, the product from the first round of nested PCR was purified. Purified first round product was extended using generic gateway primers, which incorporated a sequence encoding an N-terminal TEV protease cleavage site (Table 8.5).

Successful amplification was determined by agarose gel electrophoresis and the product from the second round of nested PCR was also purified. Cloning and sub-cloning of *neuB* was achieved through the use of the gateway system as outlined in Section 8.3. The linear product from second round nested-PCR was cloned into the donor vector pDONR-221. Ligated product was subsequently transformed into *E. coli* One Shot® TOP10 cells (Invitrogen) as outlined in Section 8.3.10. Plasmid from a selected transformant was purified and *neuB* was successfully sequence verified in the donor vector. The *neuB* gene was sub-cloned into the destination vector pDEST-17, which encodes an N-terminal His-tag.

2.2.2 Expression and purification

The *neuB* gene was transformed into three different expression cell lines to optimise the yield of soluble protein. The *neuB* gene was transformed into *E. coli* BL21 (DE3) Star cells, *E. coli* Rosetta™ 2 (DE3) cells and *E.coli* Bl21 (DE3) pBB540/pBB542 (Chaperone 3) cells. In order to compare induced against non-induced samples, two cultures of each cell line were grown overnight in LB, with one induced by IPTG and the other not. The best yield of soluble protein was attained using Chaperone 3 cells as analysed by SDS PAGE. *Nme*NANAS was thus expressed and purified from this cell line.

*Nme*NANAS was expressed and purified using the method outlined in Section 8.4. Cell pellets were resuspended in buffer and lysed via sonication. The resulting lysate was clarified by centrifugation to separate out the cell debris. Soluble protein was filtered and run through an IMAC column to purify the His-tagged protein. After the first round of IMAC purification, protein was desalted and treated with TEV protease to remove the His-tag and then run through a second IMAC column to separate the tag and TEV protease (which itself has a non-cleavable His-tag) from the target protein. SEC was used as a final, polishing step and fractions were subsequently pooled, concentrated, flash-frozen and stored at -80 °C. Samples of purified enzyme were analysed by SDS PAGE (Figure 2.3).

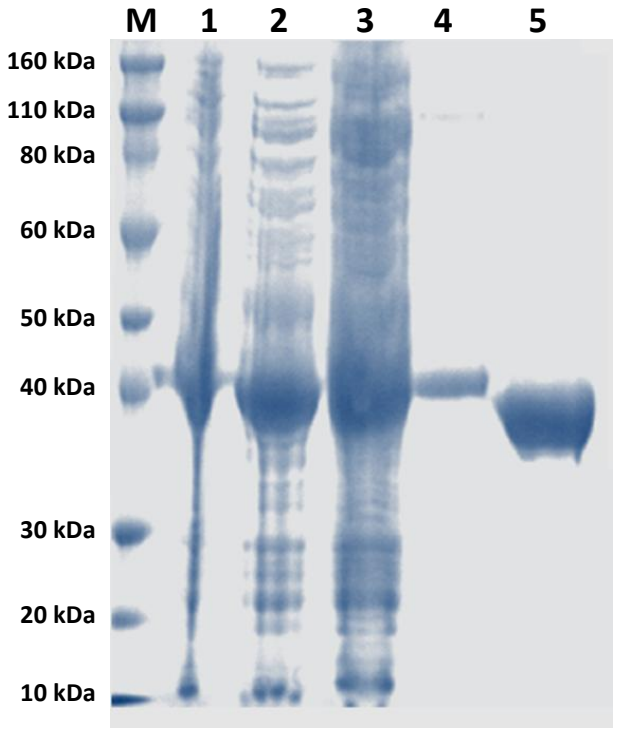


Figure 2.3 SDS-PAGE gel of *Nme*NANAS wild-type purification. Lane M contains marker with molecular weights defined. Lane 1 contains insoluble lysate, Lane 2 soluble lysate, Lane 3 first IMAC flow-through, Lane 4 first IMAC bound protein (His-tagged *Nme*NANAS) and Lane 5 purified *Nme*NANAS after TEV protease treatment, second IMAC and SEC.

2.3 Characterisation of *Nme*NANAS

2.3.1 Michaelis-Menten kinetics

Previously, two different assays have been described for determining the kinetic parameters of *Nme*NANAS. The first of these is a stopped assay, and the second is a continuous assay where NANAS activity is coupled with NANA lyase, the latter of which breaks down NANA into ManNAc and pyruvate, which is spectrophotometrically monitored (*6, 112*). NANA lyase catalyses a reversible reaction which breaks down NANA into pyruvate and ManNAc (*139, 140*). The spectroscopic assay used for kinetic analysis of *Nme*NANAS in this thesis is continuous, and directly monitors the loss of PEP. This assay has been previously described for activity assays of other PEP aldolases such as DAH7PS and KDO8PS (*123, 124*).

The kinetic parameters of wild-type *Nme*NANAS were determined using the assay described in Section 8.5.1. Cuvettes containing 50 mM BTP (pH 8), 1 mM $MnCl_2$, PEP (varied) and ManNAc (varied) were incubated for 10 min at 25 °C. The enzymatic reaction was initiated by the addition of purified enzyme. Michaelis-Menten steady-state approximations were utilised to determine kinetic parameters (Figure 2.4).

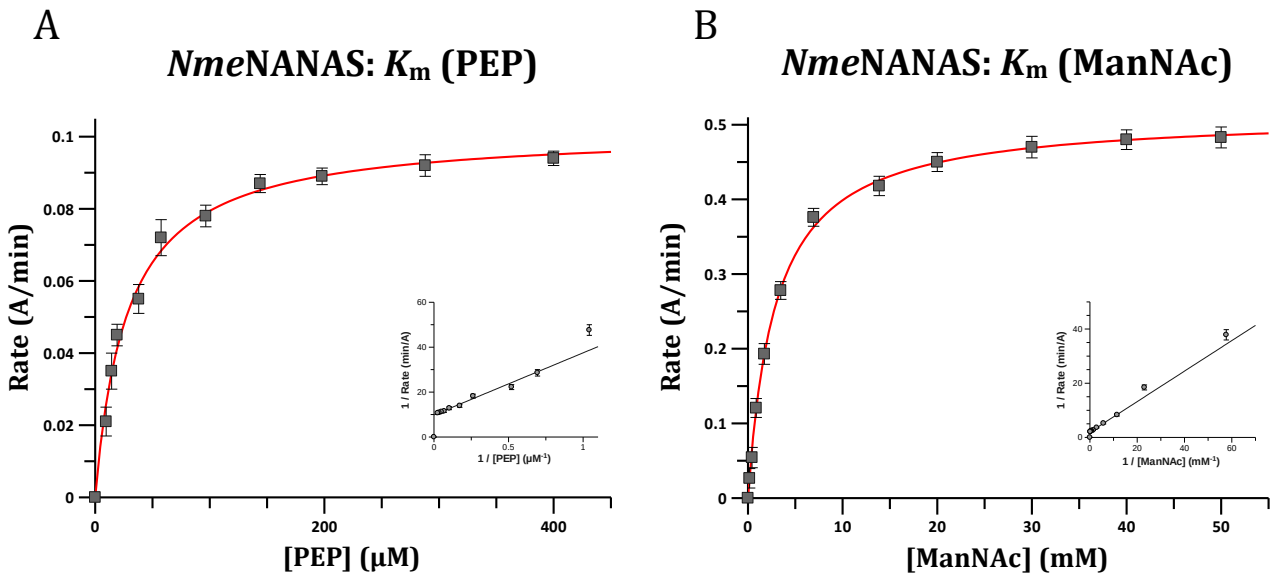


Figure 2.4 Michaelis-Menten curves for *Nme*NANAS wild-type with ManNAc and PEP. (A) *Nme*NANAS PEP K_m curve and (B) *Nme*NANAS ManNAc K_m curve. Data points are shown in black and a line of best fit in red. A double reciprocal plot is displayed as an inset for each graph.

The K_m of ManNAc for *Nme*NANAS was calculated to be 2.9 ± 0.2 mM and the K_m of PEP for *Nme*NANAS was calculated to be 28 ± 3 µM. The previously determined values for the K_m of ManNAc and PEP were 6.25 mM and 42 µM respectively (*104*). The kinetic assay used for the previously reported value was a stopped assay with thiobarbituric acid. The differences in values attained in previous studies are likely a result of variation in methodology and conditions used for analysis. The k_{cat} for the reaction was calculated to be 3.1 ± 0.1 s^{-1} with the literature value determined as 0.9 s^{-1} (*103*).

2.3.2 Mass spectrometry

The molecular mass of His_6-tagged *Nme*NANAS was analysed using mass spectrometry. A theoretical mass of 41,977 Da was calculated for *Nme*NANAS from the peptide sequence using ProtParam (*141*). The experimental mass determined by mass spectrometry was 41,977 Da which was identical to the aforementioned calculated mass.

2.3.3 Circular dichroism

CD spectroscopy was used to assess the secondary structure profile of *Nme*NANAS (Figure 2.5). The K2D3 server (*142*) was used to estimate the secondary structure composition from CD spectra. Shown in Table 2.1, estimated secondary structure is compared to values derived from the crystal structure of *Nme*NANAS (PDB code: 1XUZ).

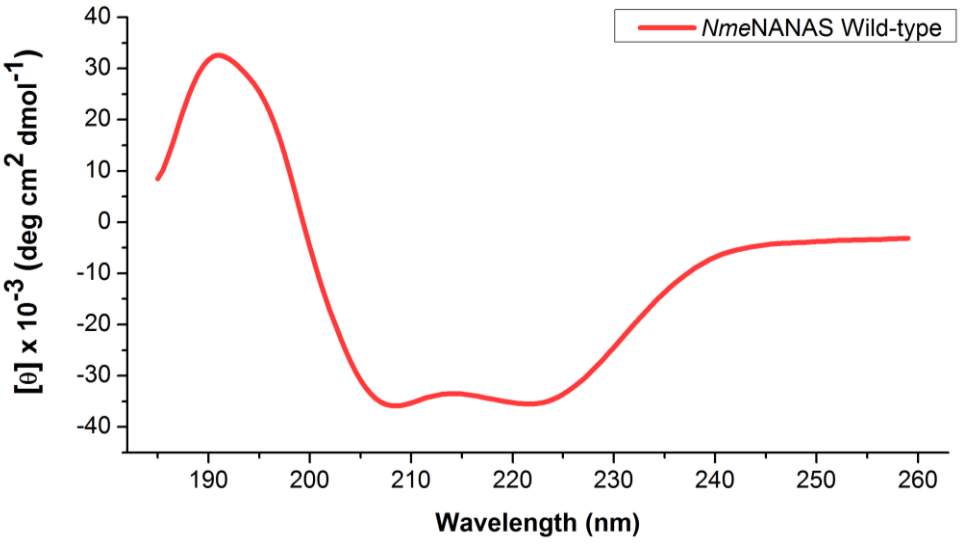


Figure 2.5 CD spectra for *Nme*NANAS wild-type (red).

	*Nme*NANAS (CD)	*Nme*NANAS (structure)
α-helix	59.9 %	50.3 %
β-sheet	12.3 %	15.5 %
other	27.8 %	34.2 %

Table 2.1 Predictions of secondary structure composition for *Nme*NANAS wild-type calculated from CD data using K2D3 server (*142*). Secondary structure composition of *Nme*NANAS structure calculated from crystallographic coordinates 1XUZ.

The results from CD reveal that the *Nme*NANAS enzyme is folded in solution (Figure 2.5). Depicted in Table 2.1, the secondary structure composition estimated from the CD data of wild-type *Nme*NANAS is comparable to the levels of α-helix and β-sheet calculated using the crystal structure *Nme*NANAS 1XUZ (*112*). The results from CD thus indicate that the solution structure of *Nme*NANAS has similar secondary structure elements to the previously elucidated crystal structure.

2.3.4 Thermal stability

Differential scanning fluorimetry (DSF) was used to assess the melting temperature of *Nme*NANAS wild-type. The method used for DSF analysis is outlined in Section 8.5.3. DSF employs a dye which increases in fluorescence when it is bound to hydrophobic protein residues. The number of previously internalised hydrophobic residues available for binding increases as a protein denatures. By steadily raising the temperature of a protein's environment and monitoring the fluorescence, the denaturation temperature can be measured as the point of greatest increase in fluorescence. Enzymes used for DSF were treated with 10 mM EDTA prior to SEC to ensure removal of trace metals. The melting point of metal-free (apoenzyme) *Nme*NANAS was determined to be 43 ± 0.5 °C.

2.4 Cloning, expression and purification of *Nme*Arg314 variants

2.4.1 Site-directed Mutagenesis

In order to assess the role of Arg314 in *Nme*NANAS, two variants of the protein were generated. In the first, Arg314 was substituted for Lys in order to diminish the hydrogen bonding capability of the side-chain, but preserve some of the functionality of the wild-type protein. For the second variant, Arg314 was substituted for Ala, removing the hydrogen bonding capability of the side-chain at this residue completely.

The *Nme*NANAS wild-type pDEST-17 vector described in Section 2.2.1 was used as a template to generate *Nme*R314A and *Nme*R314K. DNA sequence encoding an N-terminal TEV protease cleavage site and N-terminal His-tag were incorporated from the template. The methods used for PCR amplification are outlined in Section 8.3.2, and primer sequences are displayed in Table 8.6.

2.4.2 Expression and purification

The pDEST-17 plasmids bearing the variant genes for *Nme*R314A and *Nme*R314K (outlined in Section 2.4.1) were transformed into Chaperone 3 cells to optimise the yield of soluble protein. *Nme*R314A and *Nme*R314K were each expressed and purified using a similar method to that outlined for *Nme*NANAS wild-type in Section 8.4. Purified enzymes were concentrated, flash-frozen and stored at -80 °C. Samples of purified enzyme were analysed by SDS PAGE (Figure 2.6).

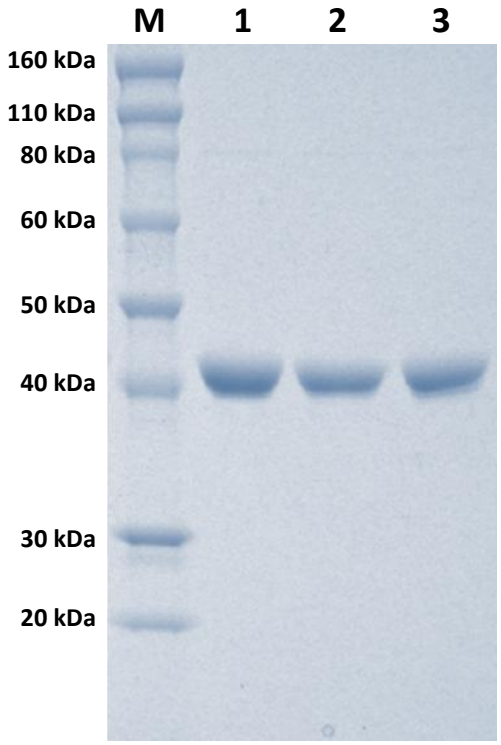


Figure 2.6 SDS-PAGE gel of purified *Nme*NANAS wild-type, *Nme*R314A and *Nme*R314K. Lane M contains marker with molecular weights defined. Lane 1 contains purified *Nme*NANAS wild-type, Lane 2 *Nme*R314A and Lane 3 *Nme*R314K.

2.5 Kinetic analysis of *Nme*Arg314 variants

2.5.1 Michaelis-Menten kinetics

To determine the kinetic parameters of the *Nme*NANAS Arg314 variants, the consumption of PEP was monitored using the assay described in Section 8.5.1. *Nme*R314K was found to be kinetically active however *Nme*R314A was inactive and no loss of PEP was observed even with excess concentrations of enzyme. Michaelis-Menten steady-state approximations were utilised to determine kinetic parameters of *Nme*R314K (Figure 2.7). Kinetic parameters were compared to the values ascertained for *Nme*NANAS wild-type (Table 2.2).

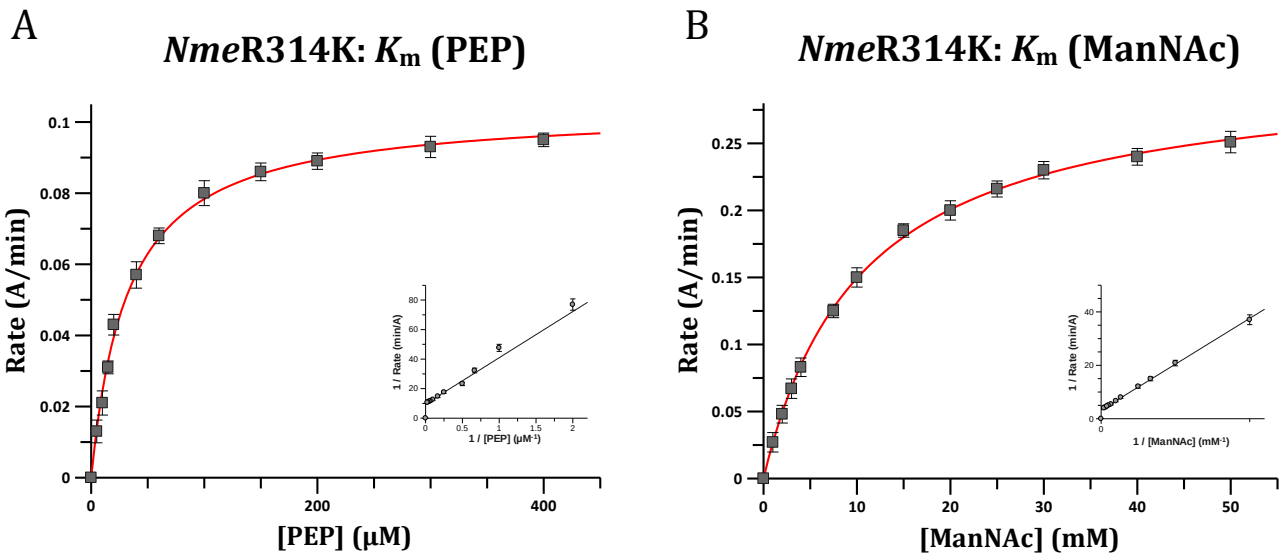


Figure 2.7 Michaelis-Menten curves for *Nme*R314K with ManNAc and PEP. (A) *Nme*R314K PEP K_m curve and (B) *Nme*R314K ManNAc K_m curve. Data points are shown in black and a line of best fit in red. A double reciprocal plot is displayed as an inset for each graph. Assays were carried out in 50 mM BTP at pH 7.5 with 1 mM $MnCl_2$, PEP (varied) and ManNAc (varied) at 25 °C. Reactions were initiated by the addition of 5 µL of purified enzyme at a stock concentration of 2.5 – 3.5 mg/mL. Michaelis-Menten steady-state approximations were utilised to determine kinetic parameters with nonlinear fitting in GraFit (Erathicus Software). To determine apparent K_m (PEP), ManNAc concentration was held at 30 mM, and the concentration of PEP varied. To determine apparent K_m (ManNAc), PEP was kept constant at 1 mM, and ManNAc concentration was varied.

	Wild-type *Nme*NANAS	R314K *Nme*NANAS	R314A *Nme*NANAS
k_{cat} (s^{-1})	3.1 ± 0.1	1.7 ± 0.1	-
K_m PEP (μM)	28 ± 3	32 ± 3	-
K_m ManNAc (mM)	2.9 ± 0.2	10.5 ± 0.1	-
k_{cat}/K_m ManNAc (s^{-1}/M)	1068 ± 108	162 ± 11	-

Table 2.2 Kinetic parameters for wild-type *Nme*NANAS, *Nme*R314K and *Nme*R314A. Values with - indicates k_{cat} of enzyme is less than 0.004 s^{-1}.

Although *Nme*R314K is catalytically active, an approximately seven-fold reduction in the specificity constant with respect to wild-type *Nme*NANAS was observed (Table 2.2). This change arises from both a reduction in the k_{cat} (approximately two-fold) and an elevation in the K_m (ManNAc) for *Nme*R314K in comparison to wild-type *Nme*NANAS. In contrast, K_m (PEP) values of the *Nme*R314K variant and wild-type *Nme*NANAS are quite similar, indicating that the PEP site is unaffected by the mutation of arginine for lysine. No loss of PEP was observed with *Nme*R314A even with a 10-fold increased addition of enzyme, indicating that this variant was unable to catalyse the aldol condensation between PEP and ManNAc.

2.6 Binding studies of *Nme*NANAS wild-type and *Nme*Arg314 variants

2.6.1 Isothermal titration calorimetry

In order to identify the causes for the loss of activity of the R314A variant of *Nme*NANAS, binding studies were undertaken using isothermal titration calorimetry (ITC). ITC was used to determine the dissociation constants (K_d) for *Nme*NANAS wild-type, *Nme*R314K and *Nme*R314A with respect to Mn^{2+}, PEP and rManNAc (Table 2.3). The Raw experimental binding curves and one-site model fitted ITC data for Mn^{2+} and PEP and rManNAc are displayed in Figure 2.8. The methods used for ITC analysis are described in Section 8.5.4.

	Wild-type	**R314K**	**R314A**
K_d **Mn^{2+} (μM)**	10.6 ± 0.5	13.5 ± 0.4	4.8 ± 0.1
K_d **PEP (μM)**	219 ± 11	406 ± 12	382 ± 26
K_d **rManNAc (μM)**	201 ± 16	206 ± 18	377 ± 26

Table 2.3 ITC derived binding constants for wild-type *Nme*NANAS, *Nme*R314K and *Nme*R314A.

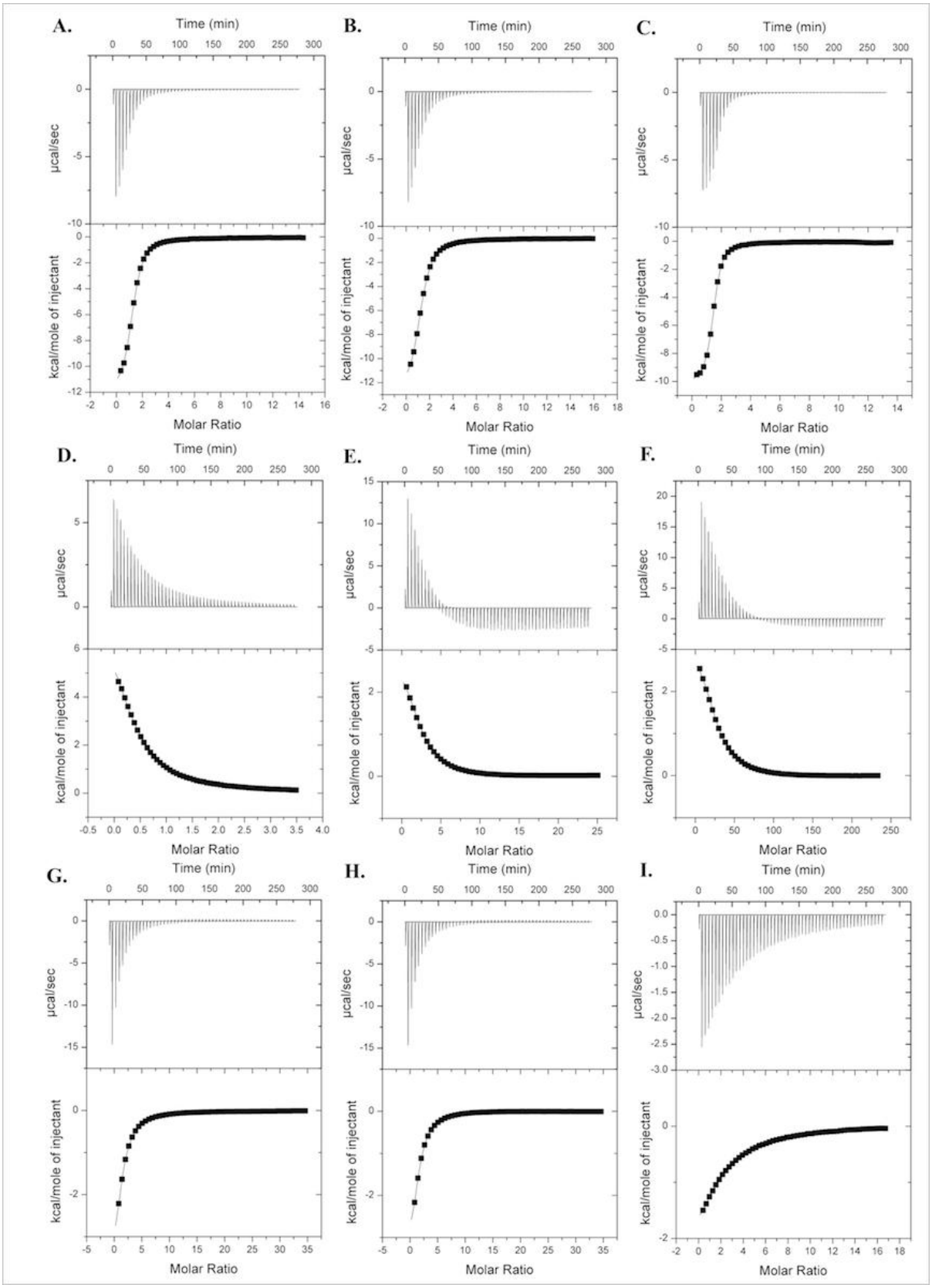


Figure 2.8 Enthalpy changes measured by isothermal titration calorimetry. (A-C) Mn^{2+} binding to metal-free *Nme*NANAS wild-type, *Nme*R314K, and *Nme*R314A respectively. (D-F) PEP binding to metal-bound *Nme*NANAS wild-type, *Nme*R314K, and *Nme*R314A respectively, (G-I) rManNAc binding to PEP/metal-bound *Nme*NANAS wild-type, *Nme*R314K and *Nme*R314A respectively. To determine $K_{d(Mn)}$, 5 mM $MnCl_2$ was titrated into 70 µM enzyme. To determine $K_{d\ (PEP)}$, 50 mM PEP was titrated into 40 µM enzyme containing a background concentration of 1 mM $MnCl_2$. To determine $K_{d\ (rManNAc)}$, 10 mM rManNAc was titrated into 100 - 200 µM enzyme containing a background concentration of 1 mM $MnCl_2$ and 1 mM PEP.

To determine the K_d values for metal with *Nme*NANAS wild-type, *Nme*R314K and *Nme*R314A, Mn^{2+} was titrated into metal-free enzyme. Mn^{2+} was chosen to investigate metal binding, as it has been previously established as the most activating metal (*103, 104*). The titration of Mn^{2+} into metal-free enzyme revealed similar affinity of these three enzymes for this metal ion. This is unsurprising given that the mutations made are not proximal to the metal binding site. Titration of PEP into metal-bound enzymes revealed that the K_d values of PEP with respect to *Nme*R314K and *Nme*R314A were almost two-fold greater than those determined for the wild-type enzyme. This correlates with the observations in Section 2.3.1 where the K_m (PEP) values of *Nme*R314K and wild-type *Nme*NANAS are noted to be similar, and indicates that the PEP site is also unaffected by mutation. Interestingly, it was noted that the binding of PEP into *Nme*NANAS wild-type, *Nme*R314K and *Nme*R314A was endothermic and hence a large favorable entropy must be associated with PEP binding (Figure 4D-F).

Titration of ManNAc into PEP-free and metal-bound wild-type or variant enzymes did not result in any observable heat changes. However, measurable heat changes were observed when rManNAc was titrated into the metal-bound enzymes with saturating levels of PEP present. Evidence for the binding of rManNAc was not observed in the absence of PEP, demonstrating that a functional binding site for the aldehydic substrate requires PEP to be bound to the protein. The dissociation constants for rManNAc with wild-type *Nme*NANAS and *Nme*R314K in the presence of metal and PEP were found to be similar, however, for the *Nme*R314A variant, the K_d for rManNAc was significantly greater.

2.7 Structural analysis of *Nme*NANAS wild-type and *Nme*Arg314 variants

2.7.1 Mass spectrometry

The masses of each of the *Nme*NANAS Arg314 variants were analysed using mass spectrometry. The experimental masses determined for both *Nme*R314K and *Nme*R314A were almost identical to those calculated from the respective peptide sequences using ProtParam (*141*). *Nme*R314K was observed to have a mass of 41,950 Da and had an expected mass of 41,949 Da, whilst *Nme*R314A was observed to have a mass of 41,890 Da and had an expected mass of 41,892 Da.

2.7.2 Thermal stability

DSF was used to assess the melting temperatures of the *Nme*NANAS Arg314 variants. The method used for DSF analysis is outlined in Section 8.5.3. Enzymes used for DSF were treated with 10 mM EDTA prior to SEC to ensure removal of trace metals. The melting points for apoenzyme *Nme*R314K and *Nme*R314A were 43 ± 0.5 °C, identical to that determined for wild-type *Nme*NANAS (Section 2.3.4). The results from DSF indicate that the thermal stability of the variant proteins have not been significantly altered through mutagenesis.

2.7.3 Circular dichroism

CD was used to assess and compare the secondary structure of the *Nme*NANAS Arg314 variants with wild-type *Nme*NANAS. Displayed in Figure 2.9, the CD profiles for both of the *Nme*NANAS Arg314 variants are almost identical to the profile for *Nme*NANAS wild-type. Furthermore, the secondary structure ratios determined from the variant CD data using the K2D3 server (*142*), were identical to the values obtained for wild-type *Nme*NANAS (Section 2.3.3). The results from CD indicate that the secondary structure of the variants is unlikely to have been significantly affected by any of the mutations made.

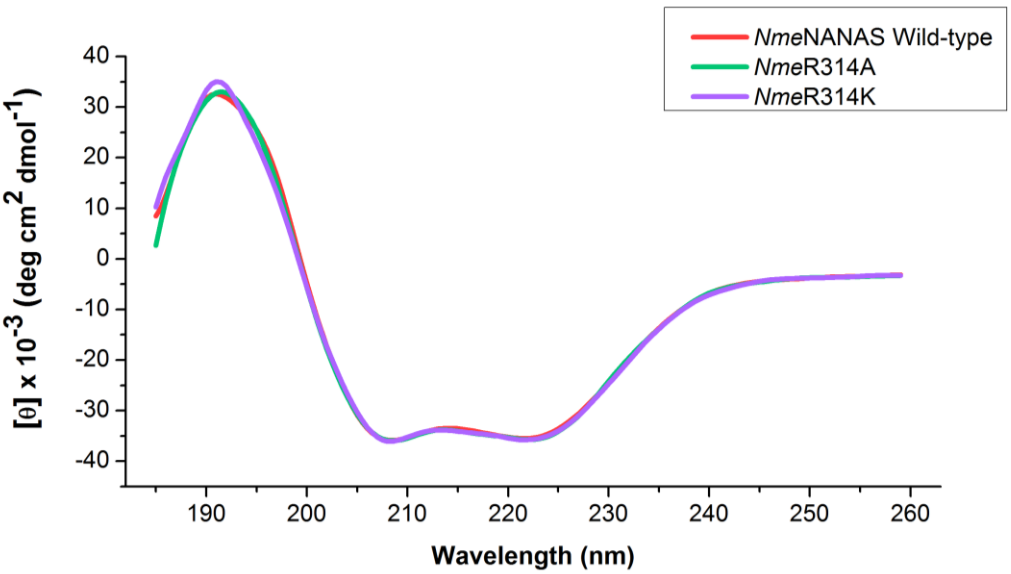


Figure 2.9 CD spectra for *Nme*NANAS wild-type (red), *Nme*R314A (green) and *Nme*R314K (purple).

2.7.4 Crystallisation

Purified *Nme*R314K and *Nme*R314A were crystallised using a modified condition previously identified for wild-type *Nme*NANAS (*112, 113*). Orthorhombic crystals of *Nme*R314K and *Nme*R314A (between 0.4 and 0.5 mm in size) were formed using the hanging drop crystallisation technique in the presence of 1.7 - 1.9 M malate and 10 mM $MnCl_2$ at pH 6.2 (Figure 2.10). Data sets were collected at the Australian Synchrotron using the MX2 beamline.

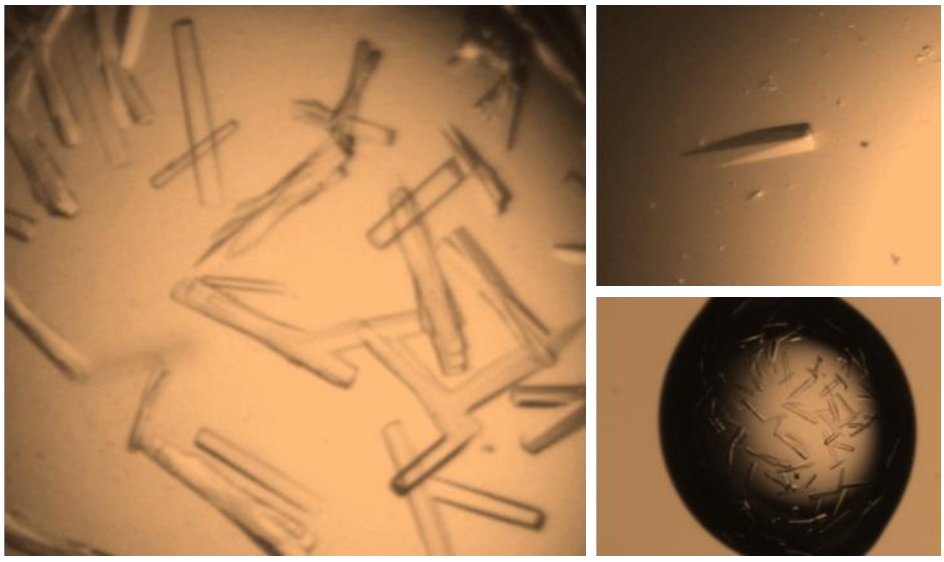


Figure 2.10 Image of orthorhombic *Nme*NANAS R314K and R314A crystals viewed through a Leica light microscope.

2.7.5 Structures of *Nme*R314A and *Nme*R314K

Crystal structures of *Nme*R314K and *Nme*R314A were successfully solved to resolutions of 2.15 and 1.75 Å respectively (Table 2.4). The overall folds for both *Nme*NANAS variants were found to be very similar to that of the malate-bound wild-type structure (PDB code 1XUU) with RMS deviations for the monomer of the malate-bound wild-type protein compared to the monomers of *Nme*R314K and *Nme*R314A, calculated as 0.165 Å (315 to 315 atoms) and 0.183 Å (319 to 319 atoms) respectively. The space group and unit cell dimensions are also retained (*112*).

	*Nme*R314K	*Nme*R314A
A. *Data collection*		
Crystal system; space group	$P2_12_12$	$P2_12_12$
Unit cell parameters (Å)	58.23, 76.05, 77.67,	58.00, 75.80, 77.52,
a, b, c, α, β, γ	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution range (outer shell) (Å)	38.00-2.15 (2.2-2.15)	27.16-1.75 (1.8-1.75)
Measurements	261,329	481,122
Unique reflections	19,392	35,205
Redundancy	13.5	13.7
Completeness (outer shell) (%)	100.0 (100.0)	100.0 (99.9)
I/σ (outer shell) (*I*)	15.9 (7.3)	21.7 (5.2)
R_{merge} (outer shell)	0.137 (0.405)	0.106 (0.508)
Wilson *B*-value ($Å^2$)	13.706	10.119
B. *Refinement*		
Resolution (Å)	2.15	1.75
R_{cryst}	0.1739	0.1613
R_{free}	0.2423	0.1972
Chain length	351	351
Observed number of residues	351	351
Water molecules	193	359
Other	2	3
Mean *B* ($Å^2$)		
Protein	23.0	17.2
Water	22.9	23.1
Other	29.4	16.9
r.m.s.d. from target values		
Bond lengths (Å)	0.0139	0.0072
Bond angles (°)	1.370	1.110
Dihedral angles (°)	5.690	5.449
Ramachandran		
Most favored (%)	97.8	98.0
Allowed (%)	2.2	2.0
Generously allowed (%)	0	0
Disallowed (%)	0	0
PDB entry	4IPJ	4IPI

Table 2.4 Crystal parameters, data collection and refinement statistics for crystal structure of *Nme*R314K and *Nme*R314A.

As aforementioned, crystals were formed in high concentrations of malic acid (1.7 - 1.9 M), and as a result bound malate was observed in the active sites of both *Nme*R314K and *Nme*R314A. Malate binding was also observed in crystals of wild-type enzyme, and soaking with high concentrations of substrate was required to displace malic acid (*112*). Crystals of both variant proteins were soaked for 24 hours with 2 M sodium phosphate (pH 6.2), 10 mM $MnCl_2$, 10 mM PEP, and 10 mM rManNAc, however, this treatment resulted in significant physical deterioration of the crystals. Additionally, a variety of crystallisation conditions and soaking conditions were screened to try to obtain crystals of the mutants in the presence of substrates PEP and ManNAc (or rManNAc). Unfortunately no additional crystallisation conditions that yielded crystals of diffraction quality were found.

Most active site residues in the variant structures are in identical positions to that for the previously elucidated wild-type structures with either malate or substrates bound (PDB codes 1XUU and IXUZ respectively). In both variant structures the backbone conformation at the site of mutation is retained and changes in density pertaining to the mutated residue are clearly apparent (Figure 2.11).

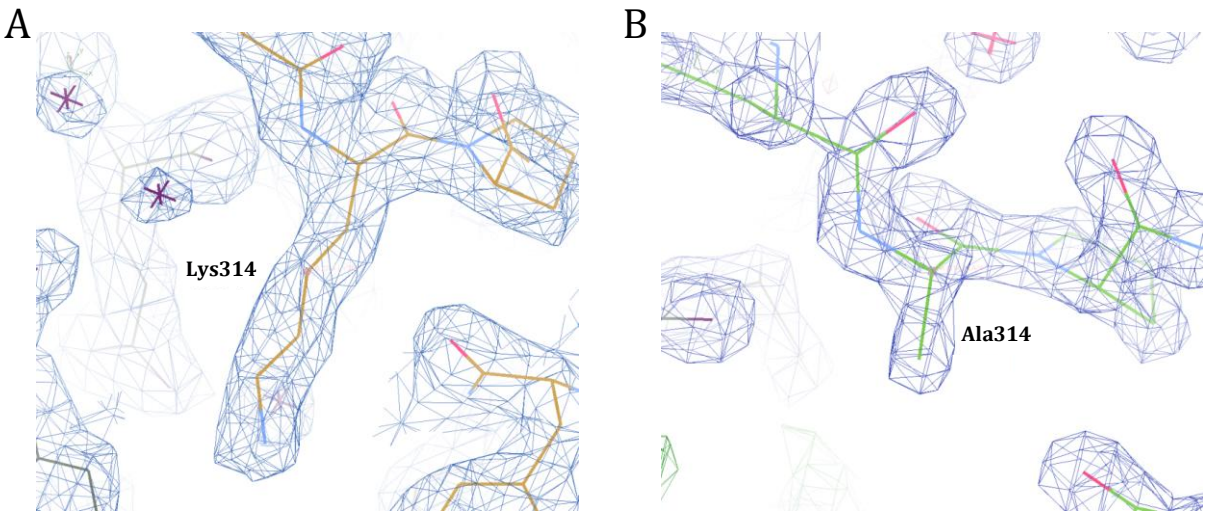


Figure 2.11 Electron density Maps of (A) *Nme*R314K and (B) *Nme*R314A showing the mutation at residue 314. The electron density map ($2F_o$-F_c) is displayed at a σ level of 1.0 in blue and the difference density map (F_o-F_c) is displayed at a σ level of 3.5 in green and red.

In the substrate-bound wild-type structure, the hydroxyl group of Tyr186 is positioned closer to the active site Mn^{2+} ion (hydroxyl is 2.8 Å away from Mn^{2+}) whereas in the malate-bound wild-type structure Tyr186 is slightly displaced (hydroxyl is 4.5 Å away from Mn^{2+}). Partial occupancy of both conformations for this residue is observed in the structures of R314K and R314A *Nme*NANAS (Figure 2.12).

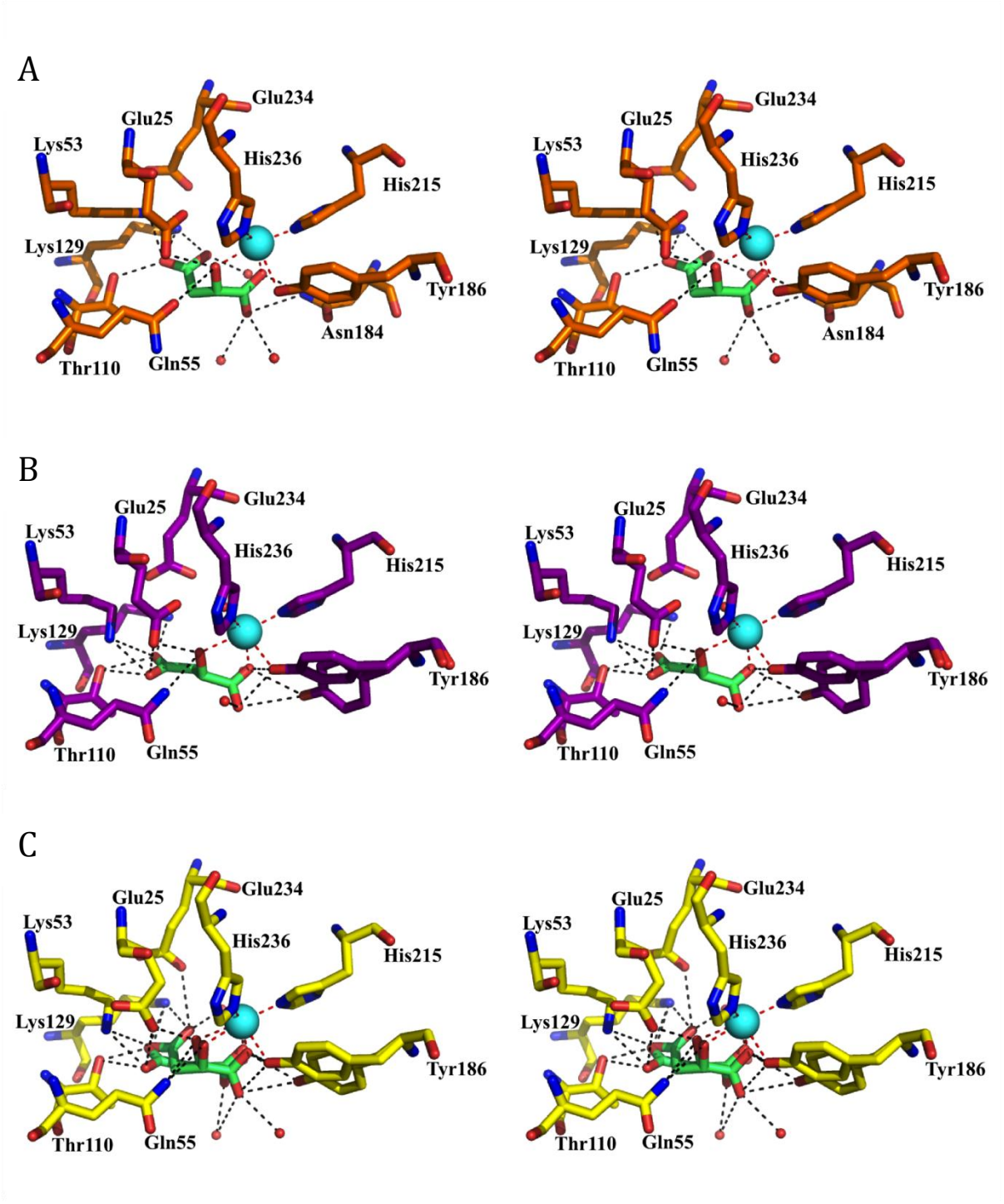


Figure 2.12 (A) Stereo view of active-site structure of malate-bound *Nme*NANAS wild-type (PDB code 1XUU), residues shown in orange. (B) Active-site structure of R314K NANAS, residues shown in purple. (C) Active-site structure of R314A NANAS, residues shown in yellow. Malate is shown in green and Mn^{2+} in blue. Hydrogen bonds to malate with distances between 2.0-3.4 Å are displayed with black dashed lines and metal-ligand interactions are displayed with red dashed lines. Note: the side-chain of Gln55 in the two mutant structures is flipped compared to that of the wild-type structure. This particular conformation was modelled in the mutant structures as it allows for more reasonable hydrogen bonds with surrounding residues.

Malate bound in the active site of R314A is observed in two alternative conformations in both enzyme variants. One of these conformations is similar to that observed in the malate bound wild-type NANAS (Figure 2.12). There is also a slight change in positioning of Mn^{2+} (1.0 Å towards Tyr186) for the malate bound structures (including the wild-type structure 1XUU) when compared to the substrate bound wild-type structure (1XUZ), which presumably arises due to the presence of different ligands in the active site. One possible consequence of the substitution of Arg314 for Ala could be the loss of the ability of the protein to exclude water molecules from the active site. A comparison between the wild-type, R314K and R314A *Nme*NANAS crystal structures around the site of mutation indicates that R314A *Nme*NANAS does not bind additional water molecules in the region where Arg314 side-chain is missing (Figure 2.13).

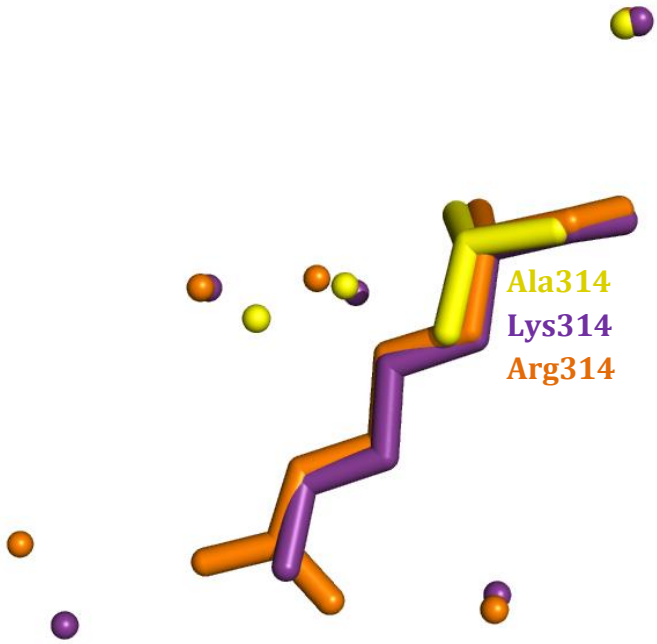


Figure 2.13 Water molecules surrounding the side-chains of residue 314. Crystal structure of the wild-type *Nme*NANAS is shown in orange (PDB code 1XUU), R314K *Nme*NANAS is purple and R314A *Nme*NANAS is shown in yellow.

2.7.6 Substrate docking in *Nme*R314A and *Nme*R314K

As discussed in Section 2.7.5, substrate-bound structures of *Nme*R314K and *Nme*R314A were unable to be obtained. In order to determine the possible substrate interactions leading to diminished activity in *Nme*R314K and loss of activity in *Nme*R314A, docking studies using the malate-bound structures of *Nme*R314K (PDB code 4IPJ) and *Nme*R314A (PDB code 4IPI) were undertaken. Docking experiments and the following analyses were carried out by Dr Wanting Jiao.

In the crystal structure of wild-type *Nme*NANAS (PDB code 1XUZ), the substrate analogue rManNAc is bound in the active site (Figure 2.2) (*112*). The C1 hydroxyl group of rManNAc that corresponds to the aldehyde of ManNAc is held in position by hydrogen bonds to Tyr186, His236 and a water molecule. The hydroxyl groups on the carbon chain of rManNAc pick up hydrogen bonds with residues Asp247, Gln55, and Asn74. The side-chain of Arg314 mainly interacts with the *N*-acetyl group of rManNAc. Molecular modelling studies were conducted to investigate the effect of R314K and R314A mutations on the binding mode of the substrate ManNAc.

ManNAc was first modelled into the wild-type active site to confirm that the correct binding pose for this substrate was reproduced in the docking calculation. The modelling study produced a total of 10 possible low energy binding poses for ManNAc in the wild-type enzyme; all of the 10 poses predicted similar positioning of the *N*-acetyl group of ManNAc as that observed for rManNAc in the crystal structure, i.e. with the *N*-acetyl group interacting with Arg314 (Figure 2.14).

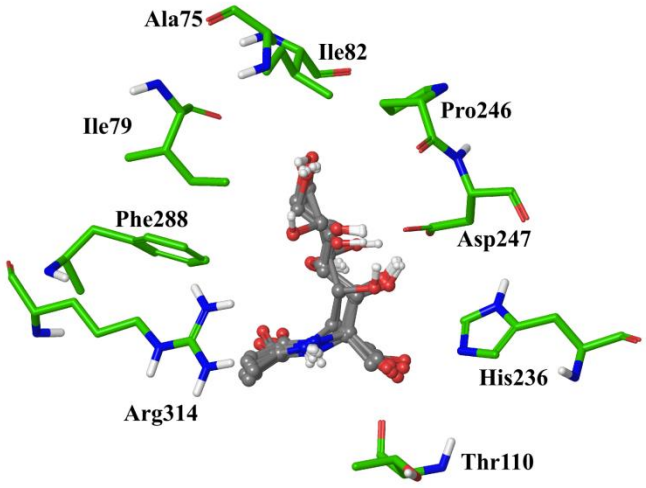


Figure 2.14 Ensembles of conformations of ManNAc in wild-type *Nme*NANAS. *Nme*NANAS receptors predicted by docking calculations using crystallographic coordinates 1XUZ. Catalytic residues are in green and ManNAc is in grey.

The distance between the *N*-acetyl carbonyl carbon atoms in the crystal structure and those in the output poses ranges from 0.15 to 0.55 Å (Table 2.5). Each of the output poses was compared to the observed binding mode of the ManNAc analogue, rManNAc, and RMS deviations were measured (Table 2.6).

Pose number	Wild-type (Å)	R314K (Å)	R314A (Å)
1	0.32	0.17	0.14
2	0.20	0.24	0.85
3	0.15	0.24	0.42
4	0.25	0.43	1.32
5	0.30	0.24	
6	0.25	0.42	
7	0.24	0.27	
8	0.25		
9	0.55		
10	0.47		

Table 2.5 Distances between the *N*-acetyl carbonyl carbon atoms in the crystal structure and those in the modelled poses.

Pose number	Wild-type (Å)	R314K (Å)	R314A (Å)
1	3.781	3.786	3.787
2	3.780	3.852	3.803
3	3.786	3.851	3.778
4	3.761	3.775	3.853
5	3.815	3.852	
6	3.849	3.777	
7	3.836	3.835	
8	3.839		
9	3.762		
10	3.799		

Table 2.6 Table of RMS deviations between output poses from modelling calculations and the observed binding pose in the crystals structure.

The output pose in the wild-type enzyme with the smallest measured RMS deviation closely resembles the binding pose of rManNAc observed in the crystal structure (Figure 2.15A, Figure 2.16A), with identical interactions between the substrate and the active site residues predicted. The aldehyde group of ManNAc in this pose is placed with a distance of 3.2 Å between the aldehyde carbon and PEP C3; and a distance of 2.7 Å between the aldehyde oxygen and the active site metal ion. The aldehyde group of ManNAc can thus coordinate to the metal ion for activation.

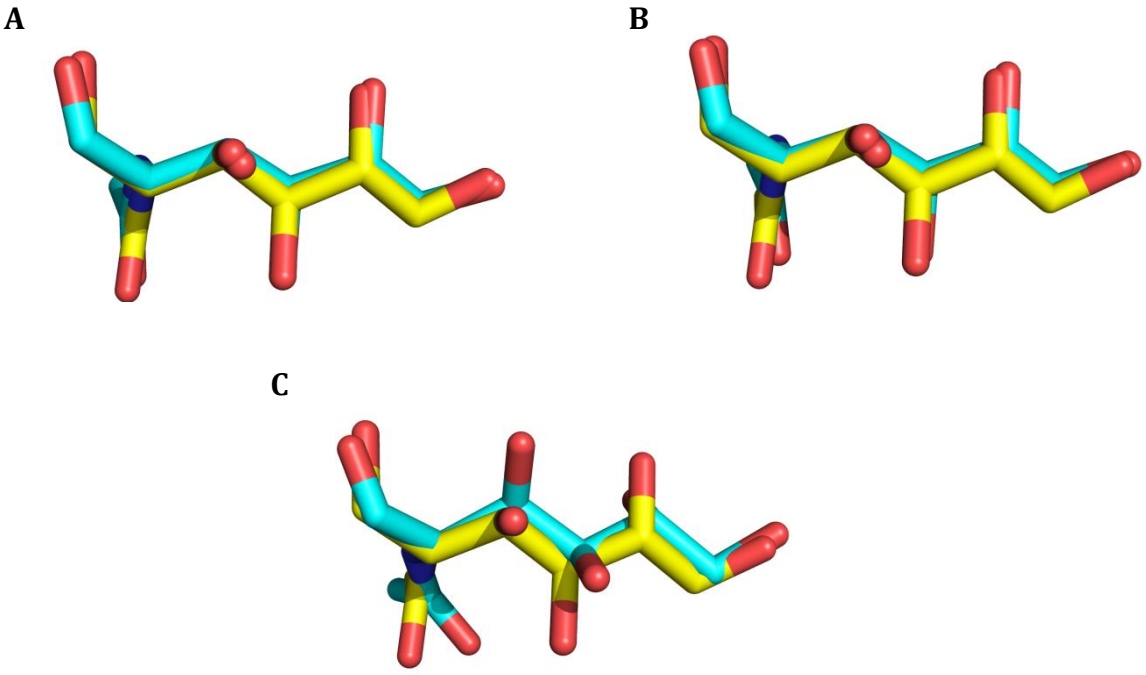


Figure 2.15 Modelled pose of ManNAc in (A) *Nme*NANAS wild-type, (B) *Nme*R314K, and (C) *Nme*R314A enzymes, superimposed with the binding position of rManNAc in the crystal structure (PDB 1XUZ). Modelled poses are in cyan whilst binding pose of rManNAc is coloured yellow.

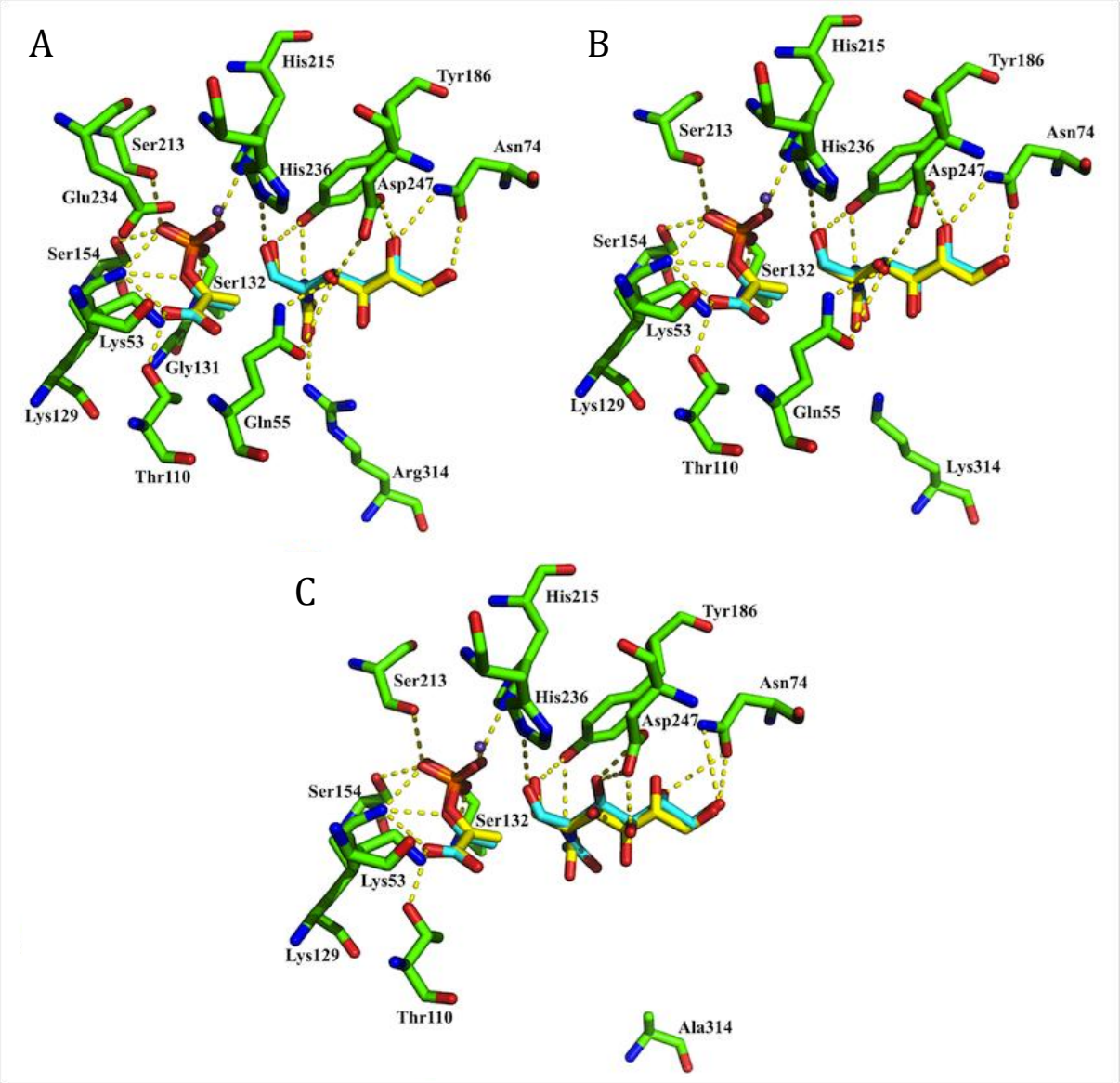


Figure 2.16 Docking poses of ManNAc in to the enzyme of (A) *Nme*NANAS wild-type, (B) *Nme*R314K, and (C) *Nme*R314A, superimposed with the binding poses of PEP and rManNAc from the crystal structure, which are displayed with yellow carbon atoms (PDB code 1XUZ). Docked ManNAc is in cyan and catalytic residues are in green. Polar interactions are depicted with yellow dashes.

The Bürgi–Dunitz angle of nucleophilic attack of PEP to the aldehyde was measured to be 94.0˚ in the 1XUZ crystal structure with rManNAc, which is slightly lower than the normal range of 100°-110° for Bürgi–Dunitz angles. This possibly arises from the presence of the tetrahedral center of rManNAc in the crystal structure. The Bürgi–Dunitz angle in the modelled pose of ManNAc with the smallest RMS deviation was measured to be 96.8°, (the range of Bürgi–Dunitz angles measured from all output poses is provided in Table 2.7). Therefore, the aldehyde group of ManNAc in this pose is held and positioned favorably for reaction to proceed.

Pose number	Wild-type (°)	R314K (°)	R314A (°)
1	97.8	98.6	98.7
2	100.4	105.9	74.3
3	98.3	105.9	88.1
4	96.8	97.4	102.8
5	68.5	105.9	
6	105.6	97.5	
7	88.6	111.7	
8	101.2		
9	85.7		
10	85.1		

Table 2.7 Table of Bürgi–Dunitz angles for output poses from modelling calculations.

Modelling of ManNAc into the *Nme*R314K variant produced 7 possible low energy binding poses for ManNAc in *Nme*R314K, and all of the 7 poses predicted similar positioning of the *N*-acetyl group as that observed in the crystal structure (Figure 2.17). The distance between the *N*-acetyl carbonyl carbon atoms in the crystal structure and those in the output poses ranges from 0.17 to 0.43 Å (Table 2.5). Following the same procedure as for the wild-type system, the output poses of ManNAc in R314K *Nme*NANAS were compared with the observed binding mode of rManNAc in the crystal structure and the RMS deviations were measured (Table 2.6).

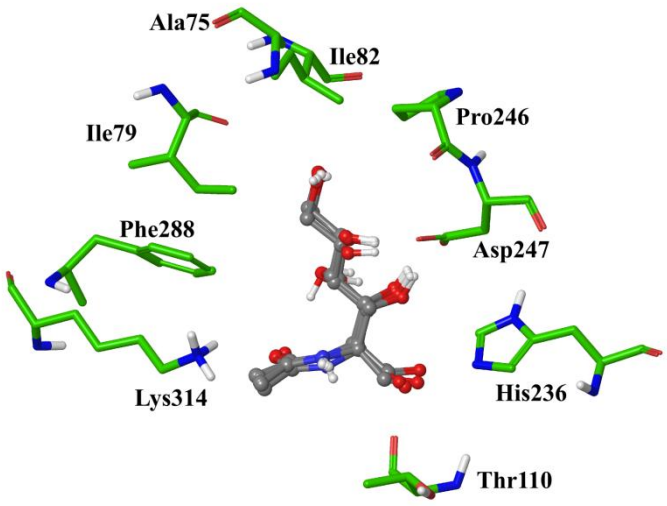


Figure 2.17 Ensembles of conformations of ManNAc in *Nme*R314K. *Nme*R314K receptors predicted by docking calculations using crystallographic coordinates 4IPJ. Catalytic residues are in green and ManNAc is in grey.

The output pose of ManNAc in *Nme*R314K with the smallest RMS deviation is still quite similar to that of rManNAc in the crystal structure (Figure 2.15B, Figure 2.16B), with all hydroxyl groups adopting the same positions as observed in *Nme*NANAS 1XUZ. The position of the *N*-acetyl group is shifted slightly away from Lys314, due to loss of the hydrogen bond caused by the Arg to Lys substitution. However, the aldehyde of ManNAc in the *Nme*R314K enzyme is still positioned similarly to that of the best pose in the wild-type enzyme, with a distance between the aldehyde carbon and PEP C3 of 3.2 Å, and a distance between the aldehyde oxygen and the active site metal ion of 2.7 Å. The Bürgi–Dunitz angle in this pose is measured to be 97.4˚. Therefore, even with slight displacement of the *N*-acetyl group of ManNAc in *Nme*R314K, the position of the ManNAc molecule (especially that of the aldehyde group) is not significantly affected, so that reaction may still proceed for the R314K variant.

The modelling of ManNAc into *Nme*R314A produced 4 poses (Figure 2.18). Although all 4 poses predicted that the *N*-acetyl groups of ManNAc occupied the same binding site, due to the loss of Arg314 side-chain, the *N*-acetyl groups show a more diverse range of positions compared to the *Nme*NANAS wild-type and *Nme*R314K variant systems (Figures 2.14 and 2.17 respectively).

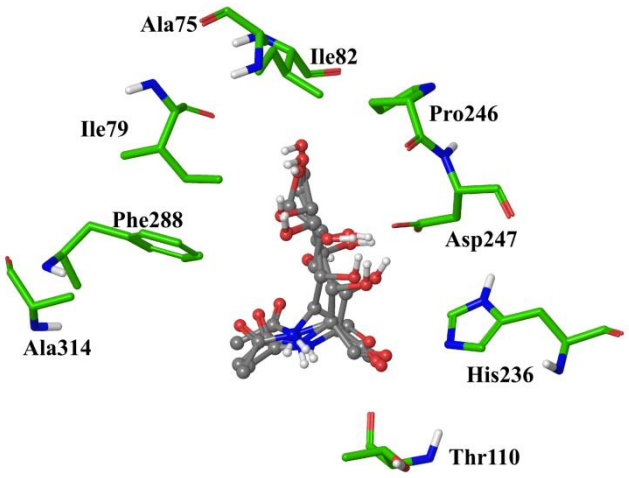


Figure 2.18 Ensembles of conformations of ManNAc in *Nme*R314A. *Nme*R314A receptors predicted by docking calculations using crystallographic coordinates 4IPI. Catalytic residues are in green and ManNAc is in grey.

The distance between the *N*-acetyl carbonyl carbon atoms in the crystal structure and those in the output poses ranges from 0.14 to 1.32 Å, in comparison to the ranges of 0.15-0.55 Å for *Nme*NANAS wild-type and 0.17-0.43 Å for *Nme*R314K (Table 2.5). The output poses were compared to the crystallographically observed binding pose (Table 2.6), and the pose in *Nme*R314A with the smallest RMS deviation appears to be very different to the poses found in wild-type *Nme*NANAS and *Nme*R314K enzymes (Figure 2.15C, Figure 2.16C), with hydroxyl groups on C3, C4 and C5 of ManNAc being placed in different positions to that observed in the crystal structure.

The R314A mutation also results in some displacement of the *N*-acetyl and aldehyde groups of ManNAc in the model. Specifically, the distance between the aldehyde carbon of ManNAc and PEP C3 is slightly increased (3.3 Å compared to 3.2 Å for *Nme*NANAS wild-type and *Nme*R314K systems), and the distance between the aldehyde oxygen and the active site metal ion was found to be similarly larger (2.8 Å compared to 2.7 Å in *Nme*NANAS wild-type and *Nme*R314K). The largest difference between the modelled pose in *Nme*R314A and that of the wild-type and *Nme*R314K systems is in the Bürgi–Dunitz angle, which is unfavourably reduced to 88.2˚(compared to 96.8˚ and 97.4˚ in wild-type and *Nme*R314K systems respectively). This observed difference in angle size is caused by the changes in conformation of ManNAc in *Nme*R314A (Table 2.7). It can be inferred that when the side-chain functionality of Arg314 is removed through mutagenesis, the enzyme has a diminished ability to correctly position the ManNAc aldehyde group and the reaction is unable to proceed.

2.8 Summary of results

Wild-type *Nme*NANAS was successfully amplified from *N. meningitidis* genomic DNA and cloned into a His-tag expression vector. *Nme*NANAS was expressed and purified by metal affinity and size exclusion chromatography. *Nme*NANAS was successfully characterised using a continuous kinetic assay previously described for DAH7PS and KDO8PS enzymes (*123, 124*). The molecular mass of *Nme*NANAS was verified using mass spectrometry and the enzyme was determined to be folded in solution using CD. The melting temperature of apoenzyme *Nme*NANAS was measured at 43 ± 0.5 °C.

In order to assess the role of residue Arg314, variant enzymes *Nme*R314K and *Nme*R314A were generated using site-directed mutagenesis. Both *Nme*R314K and *Nme*R314A were determined to be soluble and had similar CD and thermal melting profiles to wild-type *Nme*NANAS. *Nme*R314K was determined to have diminished catalytic activity whilst *Nme*R314A had no detectable activity at all.

Structures of both *Nme*R314K and *Nme*R314A were obtained using similar conditions to those previously used for wild-type *Nme*NANAS (*112, 113*). The structures of both variants contained malate in the active site due to high concentrations of the compound being present in the crystallisation conditions. Substrate bound structures were unable to be obtained even with ligand soaking and co-crystallisation.

In order to uncover the reason for loss of activity in *Nme*R314A, substrate binding studies were undertaken. Using ITC, each of the variants were determined to have similar binding affinities for both Mn^{2+} and PEP when compared with wild-type *Nme*NANAS. Interestingly, binding of ManNAc could not be observed with either *Nme*NANAS wild-type, *Nme*R314K or *Nme*R314A. Binding could be observed with rManNAc only in the presence of PEP and Mn^{2+}, indicating that PEP is required for ManNAc binding. Under these conditions, *Nme*R314A was found to bind rManNAc with a similar ITC profile to both *Nme*NANAS wild-type and *Nme*R314K.

ManNAc docking studies using the aforementioned crystal structures of *Nme*R314K and *Nme*R314A were carried out by Dr Wanting Jiao in an effort to determine the reasoning for diminished activity in *Nme*R314K and loss of activity in *Nme*R314A. The results from modelling revealed that *Nme*R314K was able to bind ManNAc in a similar fashion to the wild-type enzyme whereas *Nme*R314A was not. Several unreactive poses were observed in the docking analysis of *Nme*R314A with the angle of the aldehyde substrate relative to PEP being inappropriate for nucleophilic attack. It was thus revealed that the role of Arg314 in *Nme*NANAS is to steer the aldehyde substrate into a reactive position.

2.9 Discussion

The primary aim of the work described in this chapter was to develop an understanding of the role played by residue Arg314 in the catalytic mechanism of *Nme*NANAS. To achieve this goal, wild-type *Nme*NANAS and the variant enzymes *Nme*R314K and *Nme*R314A were successfully constructed, expressed and purified for structural and functional comparisons. Kinetic characterisation using a continuous assay revealed that *Nme*R314K had diminished catalytic activity when compared against wild-type *Nme*NANAS, whilst *Nme*R314A was found to be completely inactive. Crystal structures of both *Nme*R314K and *Nme*R314A were obtained with malate bound in the active sites.

Interestingly, the binding of ManNAc to *Nme*NANAS could not be observed by ITC, whilst binding of rManNAc could only be observed in the presence of PEP. This indicated that endothermic PEP liganding might pre-organise the binding site of ManNAc in *Nme*NANAS. The results obtained from ITC indicated that each of the variant enzymes was able to bind Mn^{2+}, PEP and rManNAc even though *Nme*R314A was inactive. ManNAc docking carried out by Dr Wanting Jiao revealed that ManNAc binding in *Nme*R314K was similar to that of wild-type *Nme*NANAS, however, modelling of ManNAc in *Nme*R314A predicted revealed several unreactive poses. In particular, the angle of the aldehyde substrate relative to PEP was not suitable for the nucleophilic reaction to occur. The combined results from kinetics, ITC and molecular modelling suggested that the primary role of Arg314 was not in binding of ManNAc but in steering the sugar substrate into a reactive position.

Chapter 3

Investigating the role of the antifreeze protein like domain from *Nme*NANAS

3.1 Background

3.1.1 Multimeric proteins

Multimeric proteins are ubiquitous in nature, and oligomerisation is widely believed to be evolutionarily favourable (*143*). Oligomerisation has many structural and functional advantages. Multimeric proteins can have enhanced stability and can often regulate accessibility and control reactivity and specificity of enzyme active sites (*144, 145*). Domain swapping is a specific case of oligomer assembly in which two or more polypeptide chains exchange units (*144*). There are many *in vivo* examples of domain swapped proteins where the interchange of domains is essential to function (*146-148*). The prevalence of multimeric proteins in nature makes understanding their evolutionary selection and function an invaluable focus of proteomics.

3.1.2 *Nme*NANAS structure and AFPL domain

The enzyme responsible for the biosynthesis of NANA is a domain swapped homodimer. The structure of *Nme*NANAS has been elucidated in the presence of Mn^{2+}, PEP and a reduced form of ManNAc (rManNAc), and has also been crystallised in the presence of a tetrahedral intermediate mimic (*112, 113*). The dimeric structure of *Nme*NANAS is displayed in Figure 3.1A. Each monomer consists of a core catalytic, $(\beta\alpha)_8$ barrel domain with a short linker extension to an AFPL domain. The AFPL domain is exclusive to the SAS family and is named for its high structural similarity to type III antifreeze proteins from fish (*103, 117*). The structure of this domain has also been elucidated by NMR from the closely related human enzyme NANA-9PS (*117*). Other members of the SAS family share high sequence homology and are also likely to contain this domain.

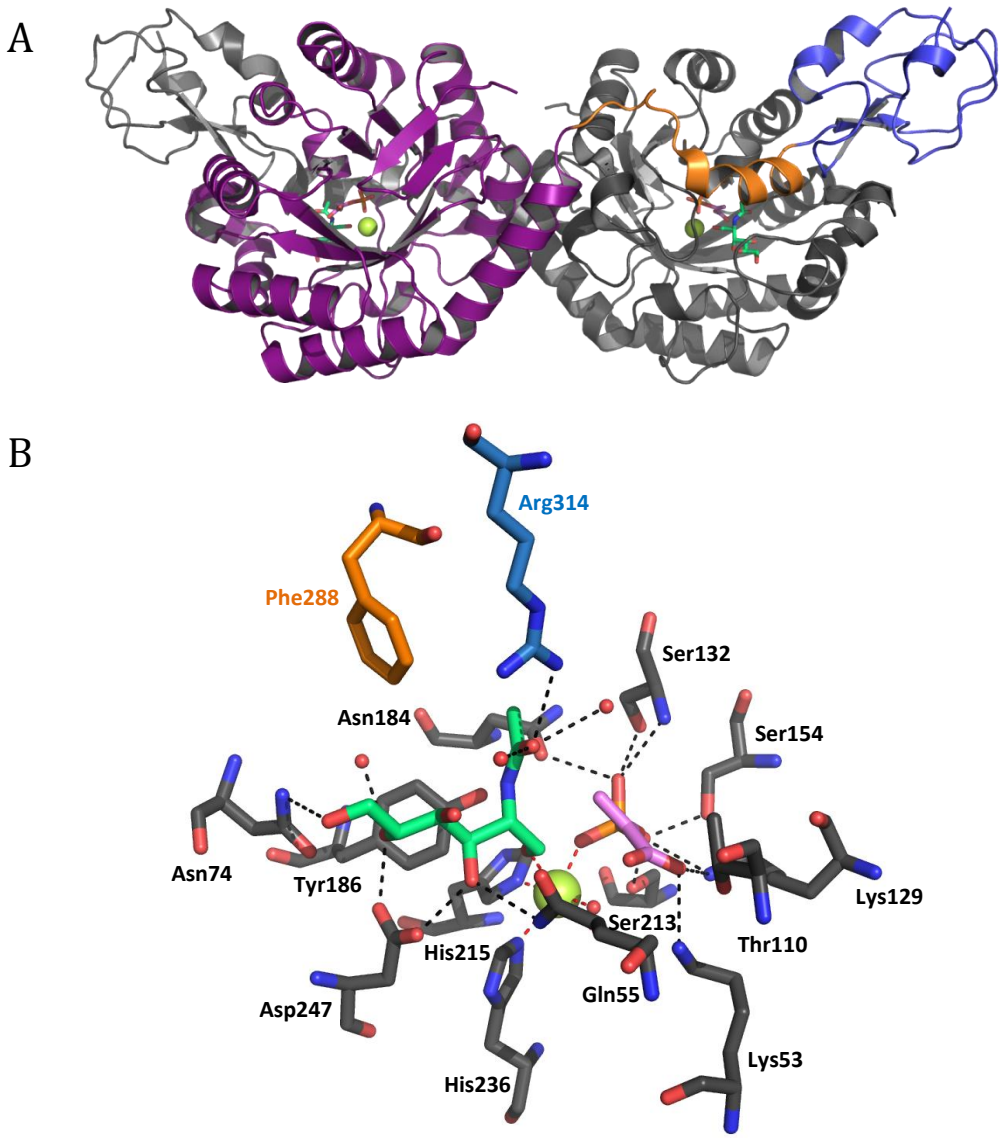


Figure 3.1 (A) *Nme*NANAS dimer with one monomer coloured grey and the second coloured by domain. The catalytic domain is shown in purple, linker domain in orange and anti-freeze protein like (AFPL) domain in blue. (B) Close up of active site. Residues contributed by the catalytic domain are in grey, linker in orange and AFPL domain in blue. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are shown in red. H-bonds are depicted with black dashes and dipolar bonds with red dashes (PDB code (1XUZ) (*112*)).

The AFPL domain caps the $(\beta\alpha)_8$ barrel domain of the opposing monomer and contributes an essential arginine into the catalytic site (Figure 3.1) (*112, 113*). Arg314 is fully conserved in AFPL domains from bacterial SASs but is not present in the AFPL domain of mammalian NANA-9-PS. Depicted in Figure 3.1B, the guanidinium functionality of Arg314 interacts with the carbonyl oxygen of the *N*-acetyl group from rManNAc. Arg314 of *Nme*NANAS has been determined to play an important role in steering ManNAc into an appropriate orientation for catalysing its reaction with PEP (Chapter 2) (*114*). While the role of Arg314 has been elucidated, the full purpose of the AFPL domain has not yet been uncovered.

SASs are closely related to both DAH7PS and KDO8PS (*122*). DAH7PS and KDO8PS catalyse analogous aldol-like condensation reactions between PEP and either E4P or A5P (*124, 149*). Each of these three PEP-utilizing aldolases shares many similarities. They each employ similar reaction chemistry (*112*), each adopt a $(\beta/\alpha)_8$ barrel catalytic domain and they are all considered to share a common evolutionary ancestor (*122*). The notable difference is that the AFPL domain is both mechanistically essential and exclusive to SASs.

3.1.3 Goals of this research

The AFPL domain is a unique C-terminal extension which is critical to the functionality of NANAS. The aim of this chapter was to investigate the role of the AFPL domain through the construction and analysis of an *Nme*NANAS truncation from which the AFPL domain has been completely excised. As discussed in Chapter 2 of this thesis, the AFPL domain contributes a critical arginine residue into the catalytic site which is involved in steering ManNAc into a reactive position. Substrate-binding analyses also indicate that PEP liganding pre-organises the active site for ManNAc binding in *Nme*NANAS (Chapter 2). The work outlined in this chapter aims to further investigate the mechanism of ManNAc binding in *Nme*NANAS through molecular dynamics and site-directed mutagenesis.

3.2 Cloning, expression and purification of *Nme*G272Term

3.2.1 Cloning of *Nme*G272Term

To assess the role of the AFPL domain, a truncated variant of *Nme*NANAS was generated which lacked the C-terminal AFPL domain (*Nme*G272Term). The truncation was made at residue Gly272 prior to the linker region (Figure 3.2).

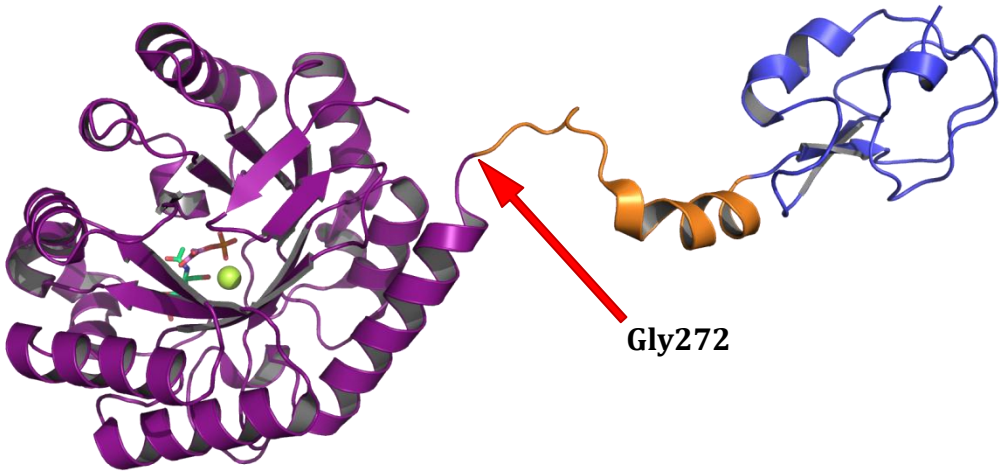


Figure 3.2 *Nme*NANAS monomer indicating position of Gly272. Catalytic domain is in purple, linker domain in orange and AFPL domain in blue. ManNAc is coloured green, PEP pink and Mn^{2+} yellow. *Nme*NANAS PDB coordinates 1XUZ (*112*).

The *Nme*NANAS wild-type pDEST-17 vector described in Chapter 2 was used as a template to generate the *Nme*G272Term construct. The *Nme*G272Term construct was created by inserting a stop codon into the *Nme*NANAS template sequence subsequent to the encoding of residue Gly272. Once synthesised, the *Nme*G272Term gene was cloned into vector pDONR 221. DNA sequence encoding an N-terminal TEV protease cleavage site was incorporated during the construction of the *Nme*G272Term gene. The *Nme*G272Term gene was sub-cloned into the expression vector pDEST-17 which encodes an N-terminal His-tag. The methods used for PCR amplification are outlined in Section 8.3.2, and primer sequences are displayed in Table 8.7.

3.2.2 Expression and purification

The vector pDEST-17 (*Nme*G272Term) was transformed into three different expression cell lines to optimise the yield of soluble protein. Vector pDEST-17 (*Nme*G272Term) was transformed into *E. coli* BL21 (DE3) Star cells, *E. coli* Rosetta™ 2 (DE3) cells and Chaperone 3 cells. In order to compare induced against non-induced samples, two cultures of each cell line were grown overnight in LB, with one induced by IPTG and the other not. The best yield of soluble protein was attained using Chaperone 3 cells as analysed by SDS PAGE. *Nme*G272Term was thus expressed and purified from this cell line. *Nme*G272Term was expressed and purified using a similar method to that used for *Nme*NANAS wild-type as outlined in Section 8.4. Purified *Nme*G272Term was concentrated, flash-frozen and stored at -80 °C. Samples of purified enzyme were analysed by SDS PAGE (Figure 3.3).

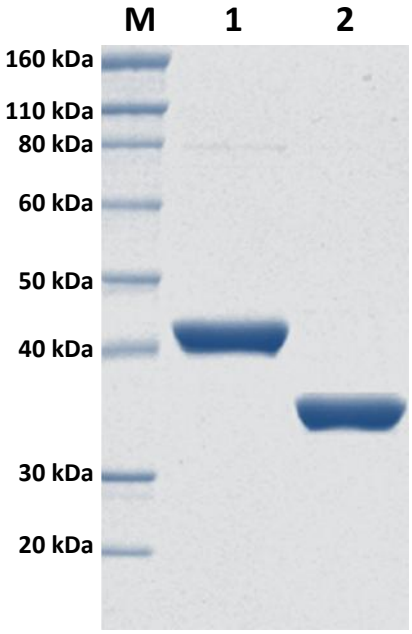


Figure 3.3 SDS-PAGE gel of purified *Nme*NANAS wild-type and *Nme*G272Term. Lane M contains marker with molecular weights defined. Lane 1 contains purified *Nme*NANAS wild-type and Lane 2 contains purified *Nme*G272Term.

3.3 Kinetic analysis of *Nme*G272Term

3.3.1 Attempted kinetic characterisation of G272Term

To determine the kinetic parameters of *Nme*G272Term, the consumption of PEP was monitored using the assay described in Section 8.5.1. *Nme*G272Term was assayed with PEP and ManNAc at several different pH levels and temperatures. No kinetic activity could be observed for *Nme*G272Term with PEP (up to 400 µM) and ManNAc (up to 500 mM) even at high protein concentrations of approximately 1 mg/mL. Lack of activity equates to a k_{cat} of enzyme less than 0.004 s^{-1}.

3.4 Binding studies of *Nme*G272Term

3.4.1 Isothermal titration calorimetry

In order to probe the causes for the loss of activity of *Nme*G272Term, binding studies were undertaken using ITC. *Nme*G272Term ligand binding was investigated with respect to metal, PEP and rManNAc using the methods described for *Nme*NANAS wild-type (Section 8.5.4). ITC data from *Nme*G272Term was compared to the previously determined dissociation constants (K_d) for *Nme*NANAS wild-type outlined in Chapter 2 (Table 3.1) (*114*). The raw experimental binding curves and one-site model fitted ITC data for Mn^{2+} and PEP with *Nme*G272Term are displayed in Figure 3.4.

	***Nme*NANAS (*114*)**	***Nme*G272Term**
K_d Mn^{2+} (µM)	10.6 ± 0.5	5.3 ± 0.1
K_d PEP (µM)	219 ± 11	320 ± 20
K_d rManNAc (µM)	201 ± 16	-

Table 3.1 ITC derived binding constants for wild-type *Nme*NANAS and *Nme*G272Term. Values with – indicate binding was not observed.

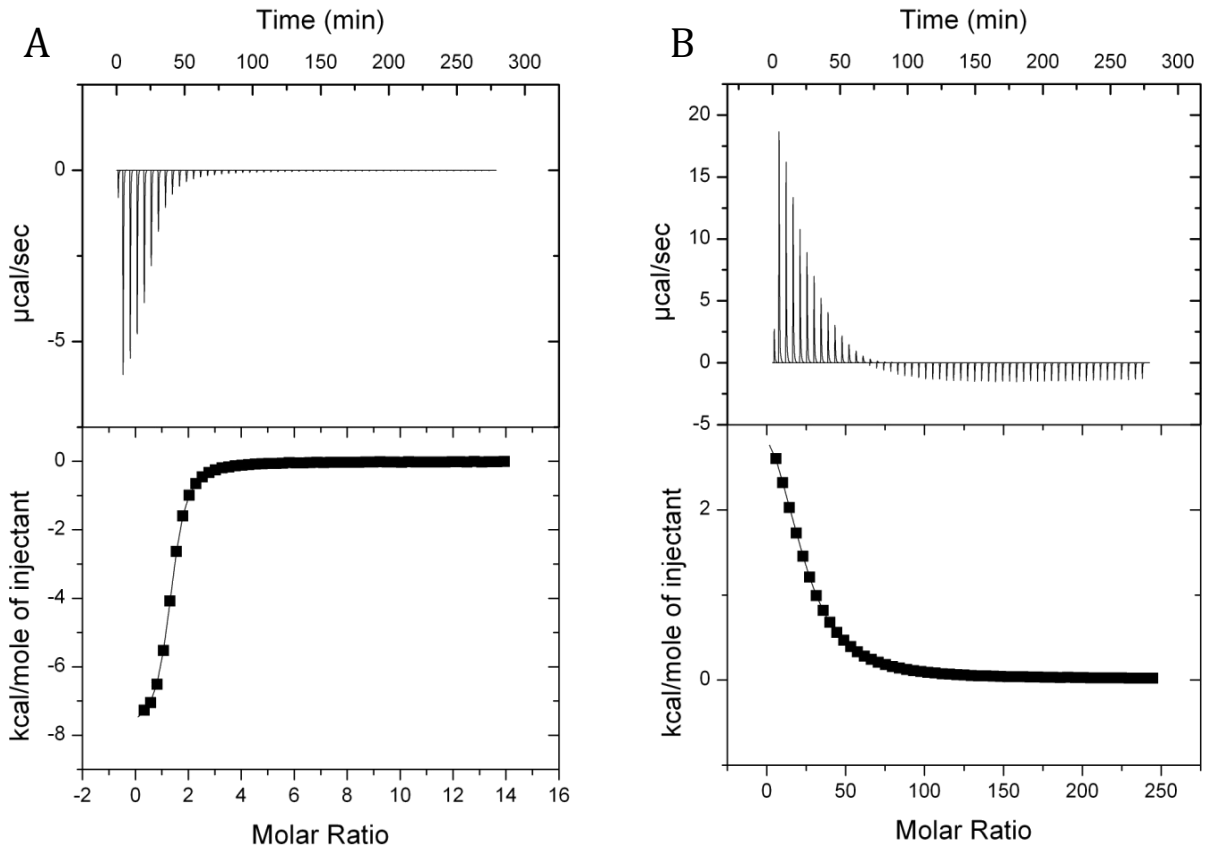


Figure 3.4 Enthalpy changes measured by isothermal titration calorimetry. (A) Mn^{2+} (5 mM) titrated into metal-free *Nme*G272Term (65 µM). (B) PEP (50 mM) titrated into *Nme*G272Term (40 µM) in the presence of Mn^{2+} (1 mM).

To determine the K_d value for metal ion with *Nme*G272Term, Mn^{2+} was titrated into metal-free enzyme. Mn^{2+} was chosen to investigate metal binding, as it has been previously established as the most activating metal for wild-type *Nme*NANAS (*5, 104*). Some differences between the Mn^{2+} and PEP affinities of *Nme*NANAS wild-type and *Nme*G272Term were noted. The affinity of *Nme*G272Term for Mn^{2+} was determined to be two-fold higher (Table 3.1). In contrast, although PEP still bound to the truncated enzyme, a decrease in the affinity of PEP for the inactive enzyme variant was apparent.

The most striking result of these studies was the change in the affinity of the truncated protein for the reduced form of ManNAc (rManNAc). As discussed in Chapter 2, this substrate analogue binds to the wild-type enzyme only in the presence of PEP (*114*), which is consistent with the proposed order of binding where PEP binds first to form the ManNAc binding site (*114*). For the truncated variant, no enthalpy of binding was observable for either ManNAc or rManNAc titrations performed either in the presence or absence of PEP and Mn^{2+}. These results indicate that loss of the AFPL domain has completely disabled the ability of *Nme*G272Term to bind ManNAc.

3.4.2 Thermal stability

DSF was used to assess the effect of substrate binding on the thermal stability of *Nme*NANAS wild-type and *Nme*G272Term. The method used for DSF analysis is outlined in Section 8.5.3. Enzymes used for DSF were treated with 10 mM EDTA prior to SEC to ensure removal of trace metals. The melting points determined for each sample were plotted graphically and are compared in Figure 3.5 as well as Table3.2.

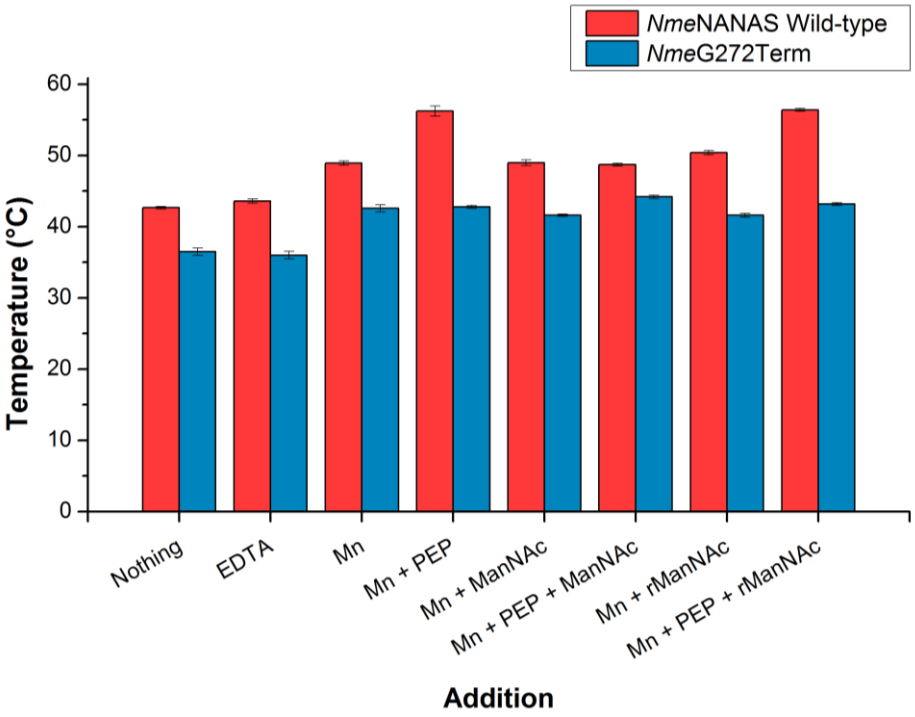


Figure 3.5 Bar graph of thermal melting points for *Nme*NANAS wild-type (red) and *Nme*G272Term (blue) in the presence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc at 10 mM. ManNAc/rManNAc was used at 1 mM when in combination with PEP.

Addition	*Nme*NANAS Wild-type (°C)	*Nme*G272Term (°C)
Nothing	42.7 ± 0.2	36.5 ± 0.5
EDTA	43.6 ± 0.3	36.0 ± 0.5
Mn^{2+}	48.9 ± 0.3	42.6 ± 0.5
Mn^{2+} + PEP	56.2 ± 0.7	42.8 ± 0.2
Mn^{2+} + ManNAc	49.0 ± 0.4	41.6 ± 0.2
Mn^{2+} + PEP + ManNAc	48.7 ± 0.2	43.2 ± 0.2
Mn^{2+} + rManNAc	50.4 ± 0.3	41.6 ± 0.3
Mn^{2+} + PEP + rManNAc	56.4 ± 0.2	43.2 ± 0.2

Table 3.2 Thermal melting points for *Nme*NANAS wild-type and *Nme*G272Term in the presence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc at 10 mM. ManNAc/rManNAc was used at 1 mM when in combination with PEP.

Apoenzyme *Nme*NANAS wild-type melted at 42.7 ± 0.2 °C, whilst apoenzyme *Nme*G272Term melted at 36.5 ± 0.5 °C. The lower thermal stability observed for *Nme*G272Term compared with wild-type enzyme is likely caused by the loss of the linker and AFPL domain. As mentioned earlier, wild-type *Nme*NANAS is a domain swapped homodimer, and the linker and AFPL domains make several contacts between the two monomer units (Section 3.1.2). The loss of these interactions may result in destabilisation of the enzyme's quaternary structure and possibly exposes more of the *Nme*G272Term barrel surface. This lower thermal stability of *Nme*G272Term when compared to wild-type *Nme*NANAS is a likely cause of the decreased melting temperature observed for the truncated enzyme.

Depicted in Figure 3.5 and Table 3.2, the addition of 1 mM EDTA to apoenzyme did not influence the thermal stability of either *Nme*NANAS wild-type or *Nme*G272Term. The absence of an effect caused by EDTA is likely due to the successful removal of trace metal ions during purification. The addition of 1 mM Mn^{2+} increased the melting points of both wild-type *Nme*NANAS and *Nme*G272Term by approximately 6 °C, to 48.9 ± 0.3 °C and 42.6 ± 0.5 °C respectively.

A significant thermal shift was observed for *Nme*NANAS wild-type in the presence of Mn^{2+} and PEP. The melting point of *Nme*NANAS with Mn^{2+} and PEP was approximately 13 °C higher than apoenzyme, with a melting temperature of 56.2 ± 0.7 °C. Interestingly this thermal shift was not observed for *Nme*G272Term which had a similar melting temperature to when only Mn^{2+} was present. *Nme*G272Term is capable of binding PEP (Section 3.4.1), however the thermal stability afforded by binding is not observed through DSF. For both *Nme*NANAS wild-type and *Nme*G272Term, the addition of 10 mM ManNAc or rManNAc in combination with 1 mM Mn^{2+} did not have any additional influence to that observed for only Mn^{2+}.

The melting temperature of wild-type *Nme*NANAS in the presence of Mn^{2+}, PEP and ManNAc was similar to that observed for *Nme*NANAS in the presence of Mn^{2+}. This is likely a result of PEP and ManNAc reacting to form NANA. Since equimolar PEP and ManNAc concentrations were used, the substrates are completely utilised and converted to product which is subsequently released by the enzyme. If this is true, only Mn^{2+} would remain bound to *Nme*NANAS, thus the melting temperature for the enzyme is similar to that observed when only Mn^{2+} is added. The similar melting temperatures for enzyme in the presence of Mn^{2+} and NANA when compared with the melting temperature for enzyme and only Mn^{2+} also indicates that *Nme*NANAS is not stabilised by the product of its reaction.

Likewise, the melting temperature for *Nme*G272Term in the presence of Mn^{2+}, PEP and ManNAc was similar to that observed for *Nme*G272Term in the presence of Mn^{2+}. Unlike the wild-type enzyme, *Nme*G272Term is unable to convert PEP and ManNAc to NANA (Section 3.3.1). The results from DSF thereby indicate that *Nme*G272Term is not stabilised by the presence of either substrate and the only stabilisation afforded is likely attributed to the presence of Mn^{2+}.

Displayed in Figure 3.5 and Table 3.2, the melting temperature for wild-type *Nme*NANAS in the presence of Mn^{2+}, PEP and rManNAc is similar to that observed for *Nme*NANAS in the presence of Mn^{2+} and PEP. As discussed in Chapter 2, *Nme*NANAS is able to bind Mn^{2+}, PEP and rManNAc. rManNAc is an unreactive analogue of ManNAc, and it is therefore likely that rManNAc, PEP and Mn^{2+} are each bound within the *Nme*NANAS active site. The presence of rManNAc does not appear to contribute any additional thermal stability over what is already afforded by PEP and Mn^{2+}.

As discussed in Section 3.4.1, *Nme*G272Term is unable to bind rManNAc. The melting temperature for *Nme*G272Term in the presence of Mn^{2+}, PEP and rManNAc is similar to that observed for *Nme*G272Term in the presence of Mn^{2+} (Figure 3.5). Similar to wild-type *Nme*NANAS, the presence of rManNAc does not confer additional thermal stability to *Nme*G272Term, even though the wild-type enzyme is known to bind rManNAc (Chapter 2).

The data obtained from DSF reveals that there is a significant thermal shift between the melting points of ligand-free and metal and PEP bound *Nme*NANAS wild-type which is not observed for *Nme*G272Term. This result indicates that PEP binding has a major effect on the thermal stability of *Nme*NANAS wild-type and that loss of the AFPL domain in *Nme*G272Term removes this property even though the *Nme*NANAS truncation appears to bind PEP relatively efficiently.

3.5 Structural analysis of *Nme*G272Term

3.5.1 Circular dichroism

CD was used to assess and compare the secondary structure of *Nme*G272Term with *Nme*NANAS (Figure 3.6). The K2D3 server (*142*) was used to estimate the secondary structure composition from CD spectra. Shown in Table 3.3, estimated secondary structure is compared to values derived from the crystal structure of *Nme*NANAS wild-type (PDB code: 1XUZ).

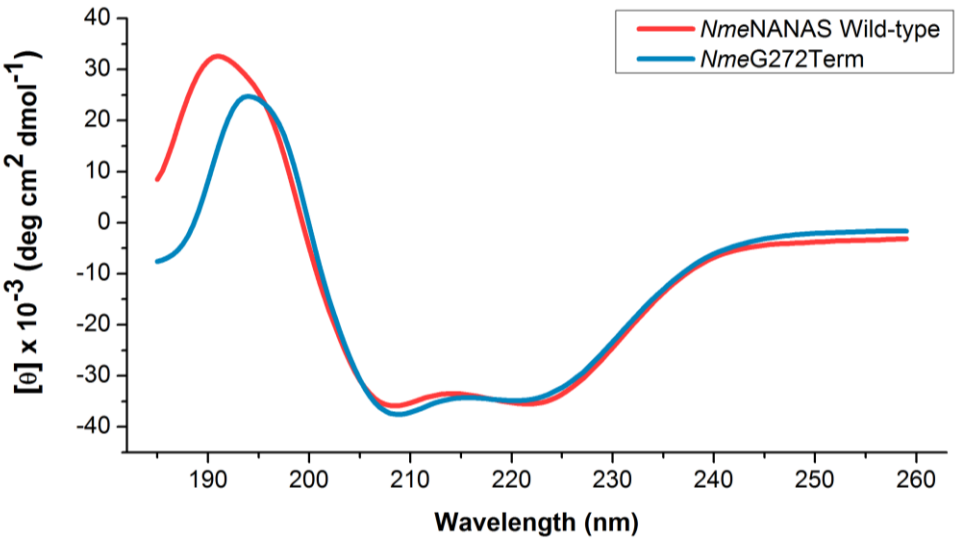


Figure 3.6 CD spectra for *Nme*NANAS wild-type (red) and *Nme*G272Term (blue).

As previously mentioned in Section 3.3, the activity of *Nme*G272Term could not be detected. It was therefore neccessary to determine that *Nme*G272Term was folded in solution using CD (Figure 3.6). Depicted in Table 3.2, the secondary structure composition calculated from the CD data of *Nme*G272Term gives comparable levels of α-helix and β-sheet to that estimated for an AFPL domain truncation derived from the crystal structure of *Nme*NANAS 1XUZ.

	*Nme*NANAS	*Nme*NANAS (structure)	G272Term	G272Term (model)
α-helix	59.9 %	50.3 %	67.6 %	73.2 %
β-sheet	12.3 %	15.5 %	21.8 %	18.0 %
other	27.8 %	34.2 %	10.6 %	8.8 %

Table 3.3 Predictions of secondary structure composition for *Nme*NANAS and *Nme*G272Term calculated from CD data using K2D3 server (*142*). Secondary structure composition of *Nme*NANAS structure and *Nme*G272Term model calculated from crystallographic coordinates 1XUZ.

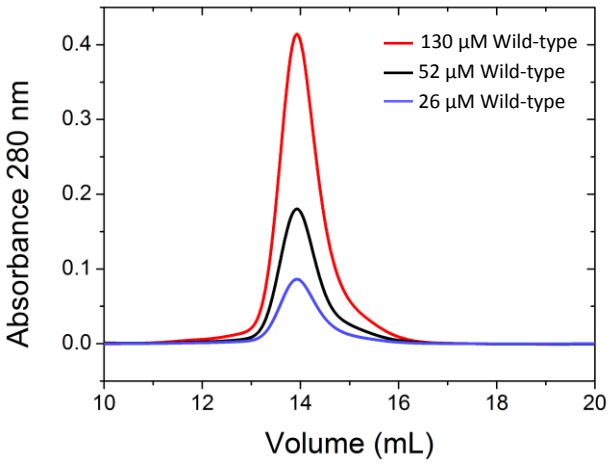
3.5.2 Mass spectrometry

The molecular mass of His_6-tagged *Nme*G272Term was analysed using mass spectrometry. A theoretical mass of 33,607 Da was calculated for *Nme*G272Term from the peptide sequence using ProtParam (*141*). The experimental mass determined by mass spectrometry was 33,608 Da which was almost identical to the aforementioned calculated mass.

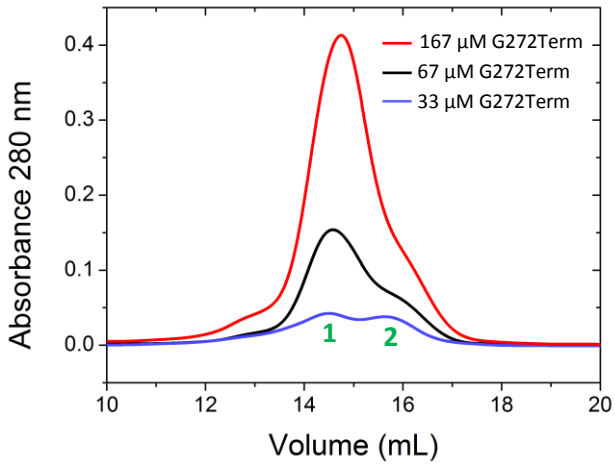
3.5.3 Analytical size exclusion chromatography

Analytical SEC was used to determine whether the removal of the linker region and AFPL domain from *Nme*NANAS had affected the quaternary structure of *Nme*G272Term. The methods used for analytical SEC are outlined in Section 8.5.6. A calibration curve was generated using protein standards of known molecular weight. The elution volume (V_e) of protein standards relative to the column void volume (V_0) was plotted against the log of the protein standards molecular mass. A linear fit of the analytical SEC calibration data was determined to have an R^2 value of 0.987. Displayed in Figure 3.7, the calibration curve was used to extrapolate the molecular masses of *Nme*NANAS wild-type and *Nme*G272Term.

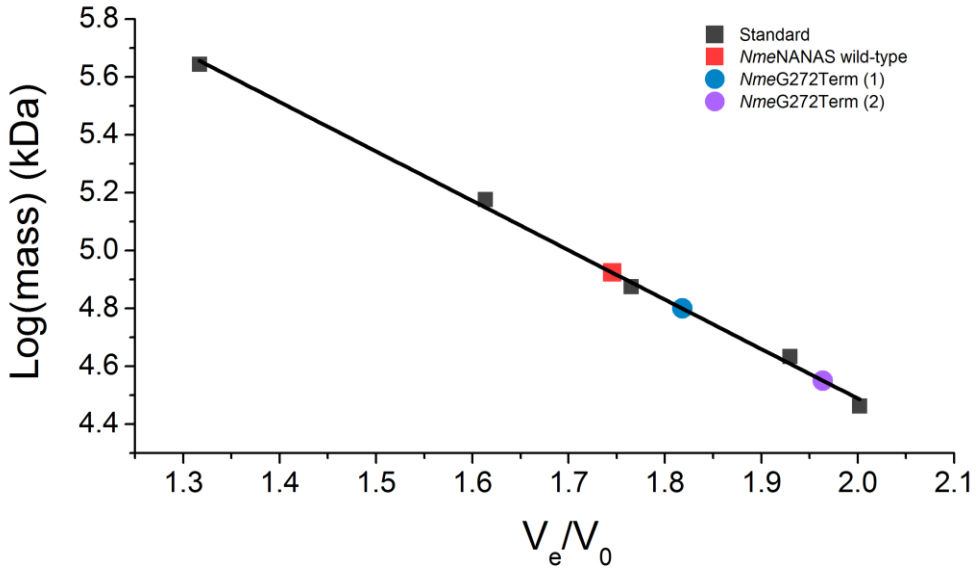
A



B



C



D

	***Nme*NANAS**	***Nme*G272Term (1)**	***Nme*G272Term (2)**
V_e/V_0	1.74	1.82	1.96
Estimated mass (kDa)	81	61	34
Theoretical mass of dimer (kDa)	77	60	30

Figure 3.7 Analytical size exclusion chromatograms measured at 280 nm absorbance. (A) Chromatography traces of *Nme*NANAS wild-type at 130 µM in red, 52 µM in black and 26 µM in blue. (B) Chromatography traces of *Nme*NANAS G272Term at 167 µM in red, 67 µM in black and 33 µM in blue. Monomer peak indicated as 1 and Dimer peak indicated as 2. (C) Analytical SEC calibration curve. Protein standards are marked in grey, *Nme*NANAS wild-type in red, *Nme*NANAS G272Term peak 1 (first elution) in blue and peak 2 (second elution) in purple. A line of best fit is plotted in black. (D) Table of estimated and theoretical masses for *Nme*NANAS and *Nme*NANAS G272Term.

As shown in Figure 3.7, wild-type *Nme*NANAS eluted with a single peak for all protein concentrations, consistent with the dimeric form of the enzyme. The estimated weight for dimeric *Nme*NANAS wild-type was 81 kDa with a calculated weight of 77 kDa.

In contrast two peaks were observed for *Nme*G272Term, with the higher retention time peak increasing in relative intensity as the protein concentration was decreased (Figure 3.7). These two peaks have the expected retention times of a dimeric and monomeric form of the *Nme*G272Term protein. The estimated weight for *Nme*G272Term peak 1 was 61 kDa with a calculated weight of 60 kDa for dimeric protein. Estimated weight for *Nme*G272Term peak 2 was 34 kDa with a calculated weight of 30 kDa for monomeric protein.

The results from analytical SEC indicate that wild-type *Nme*NANAS is a dimer at the concentrations tested. Furthermore, quaternary stability is decreased in the AFPL domain truncation resulting in a concentration dependent, monomer-dimer equilibrium. This reveals that the AFPL domain plays an important role in stabilising the quaternary structure of *Nme*NANAS.

3.5.4 Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC) was also used to assess the quaternary structures of wild-type *Nme*NANAS and *Nme*G272Term. To determine the stability of their quaternary states, sedimentation velocity experiments were caried out at varied protein concentrations. To determine whether the quaternary states of wild-type *Nme*NANAS and *Nme*G272Term were concentration dependent, samples were run at 0.5 mg/mL, 1.0 mg/mL and 2.0 mg/mL. Displayed in Figure 3.8 and 3.9, data was fitted to a continuous size-distribution (c_s) model using the program SEDFIT (*150, 151*). Weight average molecular masses were calculated using weight average sedimentation coefficients and f/f_0 values from c_s fits.

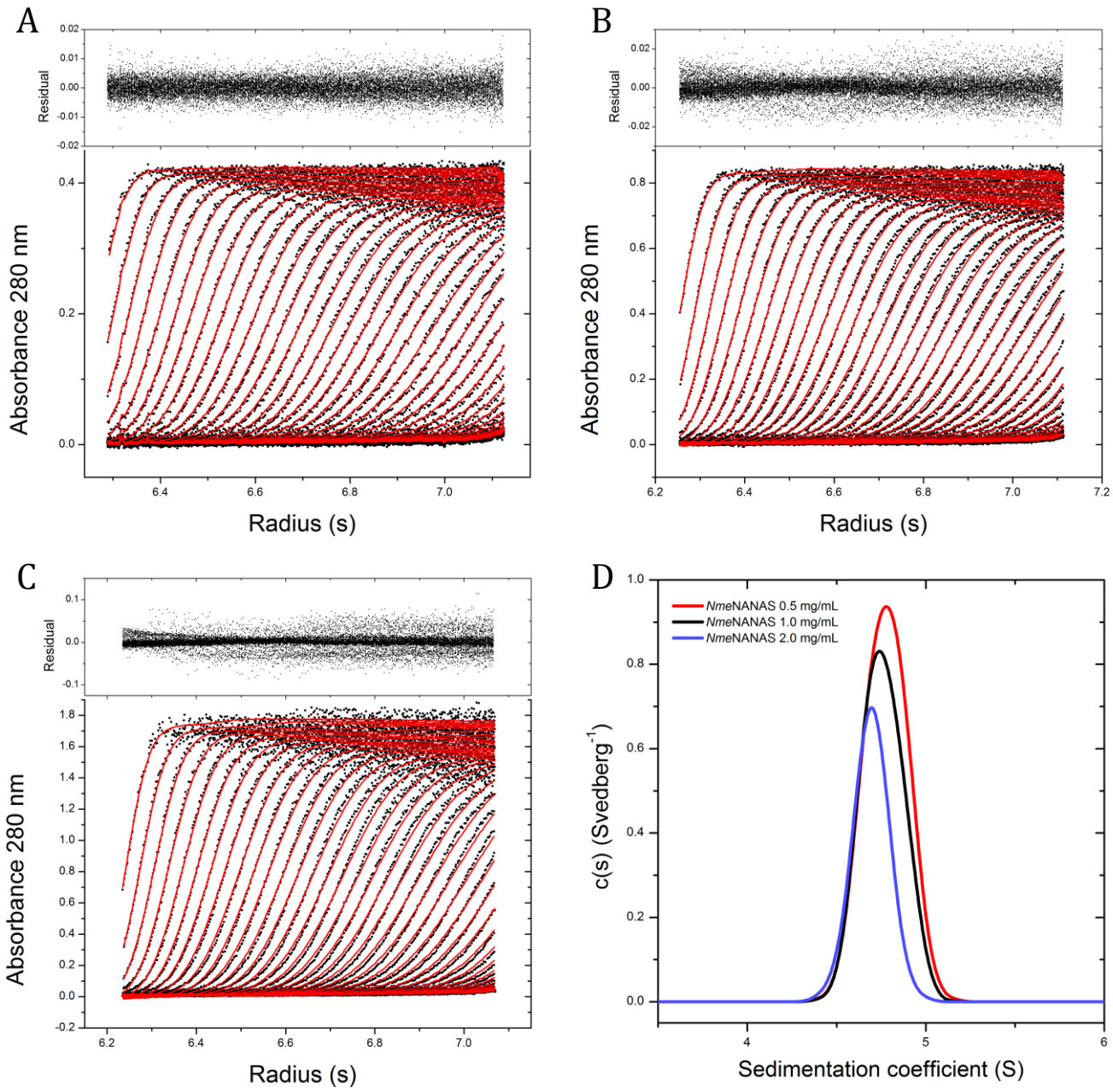


Figure 3.8 Analytical ultra centrifugation fits measured at 280 nm absorbance. (A) *Nme*NANAS wild-type at 0.5 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (B) *Nme*NANAS wild-type at 1.0 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (C) *Nme*NANAS wild-type at 2.0 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (D) Normalised c(s) profiles for *Nme*NANAS wild-type at 0.5 mg/mL (red), 1.0 mg/mL (black) and 2.0 mg/mL (purple).

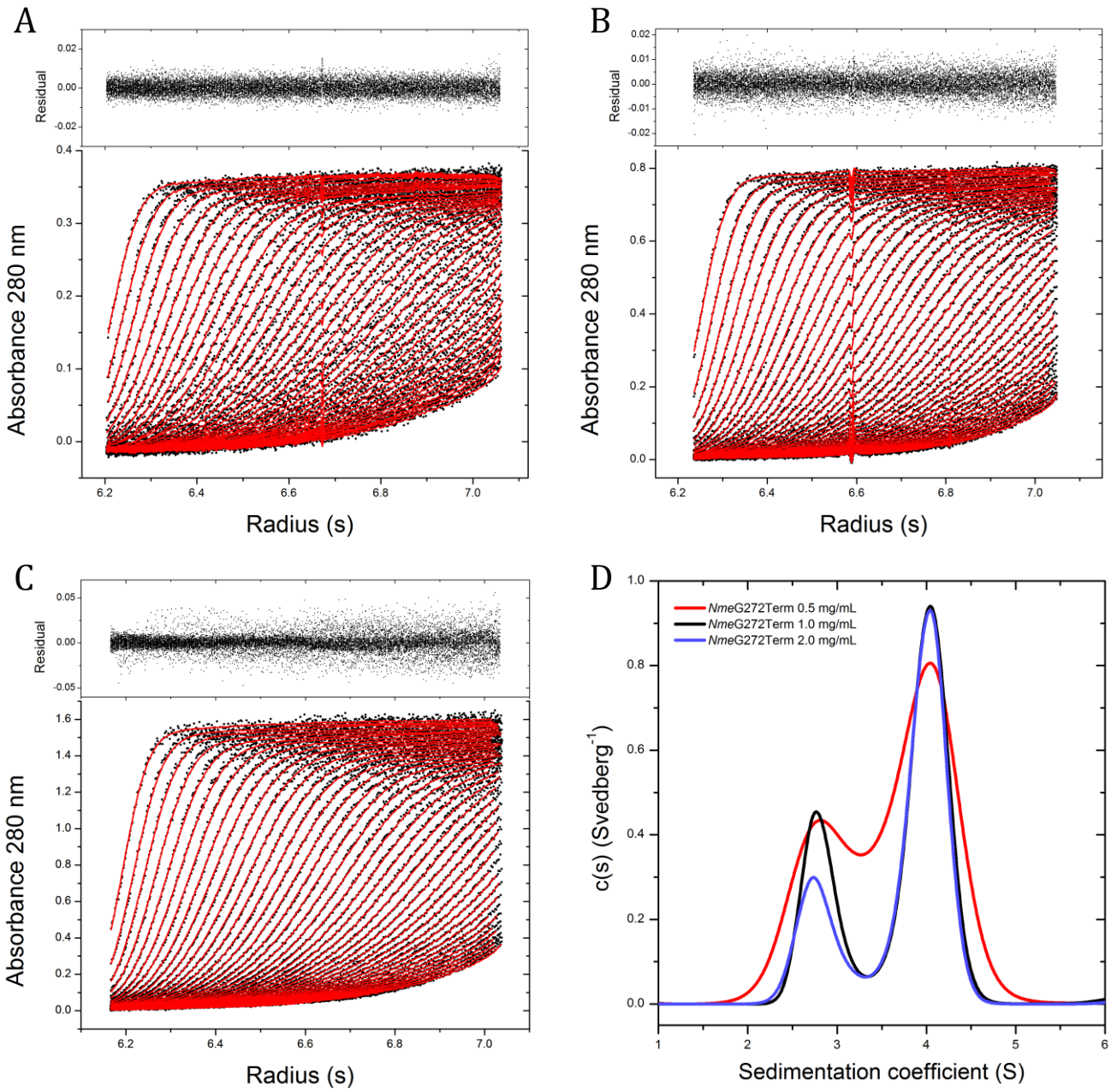


Figure 3.9 Analytical ultra centrifugation fits measured at 280 nm absorbance. (A) *Nme*G272Term at 0.5 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (B) *Nme*G272Term at 1.0 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (C) *Nme*G272Term at 2.0 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (D) Normalised c(s) profiles for *Nme*G272Term at 0.5 mg/mL (red), 1.0 mg/mL (black) and 2.0 mg/mL (purple).

Wild-type *Nme*NANAS was determined to be a single species at each of the tested concentrations (Figure 3.8). The estimated molecular weight from AUC for *Nme*NANAS wild-type was 68 kDa, which is approximately that of dimeric *Nme*NANAS. In contrast, *Nme*G272Term was present as two species (Figure 3.9). One peak corresponded to an estimated molecular weight of 31 kDa (close to the 30 kDa molecular weight calculated for monomer *Nme*G272Term) and a second peak to 51 kDa (close to the 60 kDa molecular weight calculated for dimer *Nme*G272Term). The equilibrium between these two species was demonstrated by the change in relative proportions of these two species with different protein concentrations; the ratio of monomer peak intensity to dimer peak intensity increased as the concentration of enzyme was decreased.

The results obtained from AUC correlate with analytical SEC data for *Nme*NANAS wild-type and *Nme*G272Term (Section 3.5.3). Whilst *Nme*NANAS is a dimer at each of the concentrations analysed by AUC and analytical SEC, both techniques reveal a schizophrenic quaternary state in *Nme*G272Term. In addition, the observations of a destabilised quaternary structure correlate with the decreased thermal stability observed in *Nme*G272Term when compared with *Nme*NANAS wild-type (Section 3.4.2). Each of the aforementioned analyses indicates that the AFPL domain plays an important role in quaternary structure stabilisation of *Nme*NANAS.

3.5.5 Attempted crystallisation

Several attempts at crystallising *Nme*G272Term were made. *Nme*G272Term was screened using conditions modified from those previously described for crystallising wild-type *Nme*NANAS (*112*). Protein crystallisation conditions were also screened using commercially available Morpheus, PACT and JCSG screens. Protein was concentrated to 5 mg/mL, 10 mg/mL and 15 mg/mL and screened in the absence and presence of ligands. Screens were incubated at 5 °C and 20 °C for several months, but none of the conditions that were screened yielded crystals of diffraction quality. Crystallisation of *Nme*G272Term is likely to have been problematic due to the unstable quaternary state of the protein (Sections 3.5.3 and 3.5.4). Low protein concentrations at which *Nme*G272Term would be predominantly monomeric resulted in rapid precipitation of protein, and thus screening for crystallisation conditions at these concentrations were not feasible.

3.5.6 Surface interactions of *Nme*NANAS wild-type and *Nme*G272Term

As outlined in Sections 3.5.3 and 3.5.4, the *Nme*NANAS AFPL domain truncation *Nme*G272Term has a concentration dependent quaternary state which is not observed for the wild-type enzyme. The AFPL domain is therefore involved in stabilising the dimeric form of *Nme*NANAS. In order to assess the differences in intermolecular interactions between the dimeric form of wild-type *Nme*NANAS and *Nme*G272Term, their structures were investigated using PISA (*152*). The crystal structure coordinates 1XUZ were used to assess wild-type *Nme*NANAS (*112*). Since a crystal structure of *Nme*G272Term could not be obtained, an AFPL domain truncation of the *Nme*NANAS 1XUZ structure was used to approximate the structure of *Nme*G272Term for PISA analysis.

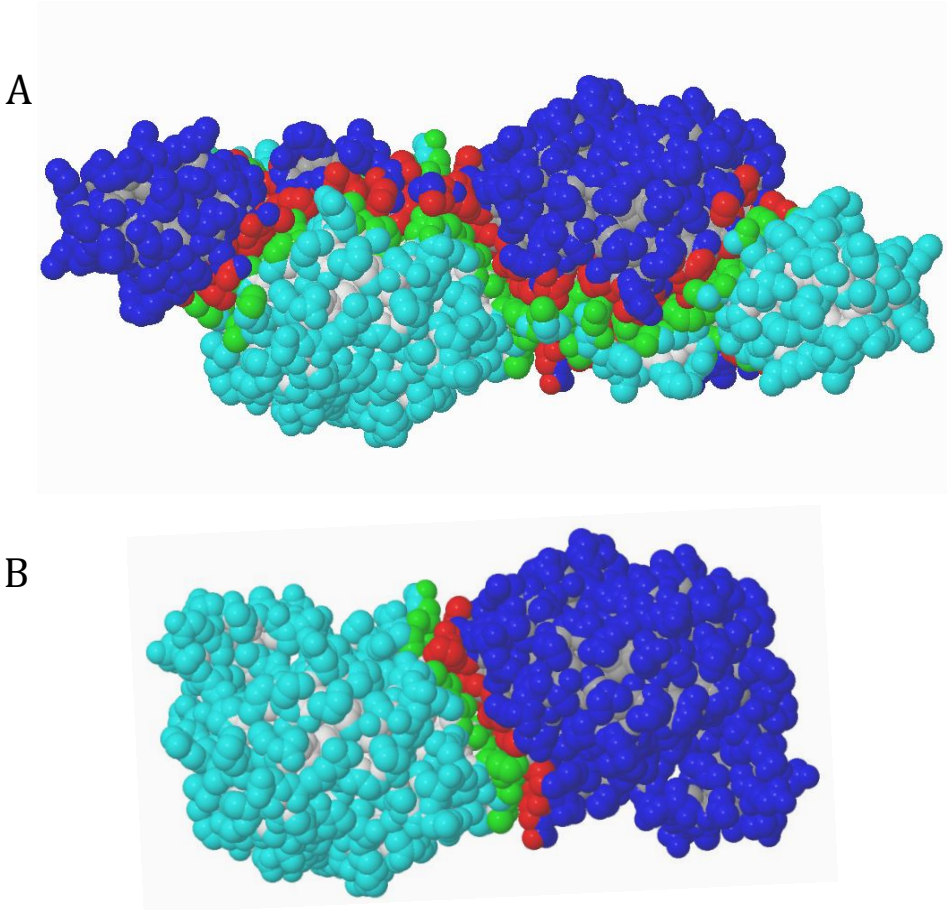


Figure 3.10 Intermolecular interactions of dimeric *Nme*NANAS wild-type (A) and *Nme*G272Term (B) as determined by PISA (*152*). Chain A of *Nme*NANAS wild-type and *Nme*G272Term are coloured cyan and Chain B in dark blue. Chain A surface interactions are in green and Chain B surface interactions are in red. *Nme*NANAS structure and *Nme*G272Term model were derived from crystallographic coordinates 1XUZ (*112*).

Depicted in Figure 3.10, a significant percentage of the intermolecular contacts present in *Nme*NANAS are lost in *Nme*G272Term. As calculated by PISA (*152*), 31.8 % of the barrel surface is involved in intermolecular contact for *Nme*NANAS wild-type whereas in comparison, only 7.1 % of the barrel surface for *Nme*G272Term would make contact in its dimeric form. Thus 77.6 % of the intermolecular contact between the monomers in *Nme*NANAS wild-type is contributed by the AFPL and linker domain. The number of interfacing atoms ($^{i}N_{at}$), was calculated from PISA to be 395 in *Nme*NANAS wild-type and only 86 in *Nme*G272Term. This significant loss of interdomain interaction is likely to be the causative factor for instability observed with dimeric *Nme*G272Term (Sections 3.5.3 and 3.5.4).

Additionally, the solvation free energy gain yielded upon interface formation ($\Delta^{i}G$), was determined to be -44.0 kcal/mol for wild-type *Nme*NANAS and -15.1 kcal/mol for *Nme*G272Term. A negative $\Delta^{i}G$ value indicates hydrophobic interface contact or positive protein affinity but does not include the effects of satisfied hydrogen bonds and salt bridges across the interface (*152*). The values derived from PISA thus indicate that the dimerisation of wild-type *Nme*NANAS is more energetically favourable than that for *Nme*G272Term.

3.5.7 Small-angle X-ray scattering

Small angle X-ray scattering (SAXS) profiles were determined to assess the solution structures of *Nme*NANAS wild-type and *Nme*G272Term and to probe conformational changes of the proteins upon ligand binding. The SAXS profiles of *Nme*NANAS wild-type and *Nme*G272Term were determined using the methods outlined in Sections 8.5.8 and 8.5.9.

Displayed in Figure 3.11, the scattering profile of wild-type *Nme*NANAS without ligand present is compared to the scattering profile of *Nme*NANAS in the presence of Mn^{2+} and PEP. The experimental scattering obtained for *Nme*NANAS is also compared with CRYSOL (*153*) fits generated from the ligand bound *Nme*NANAS crystal structure 1XUZ (*112*). The structural parameters calculated for *Nme*NANAS via SAXS, are compiled in Table 3.4.

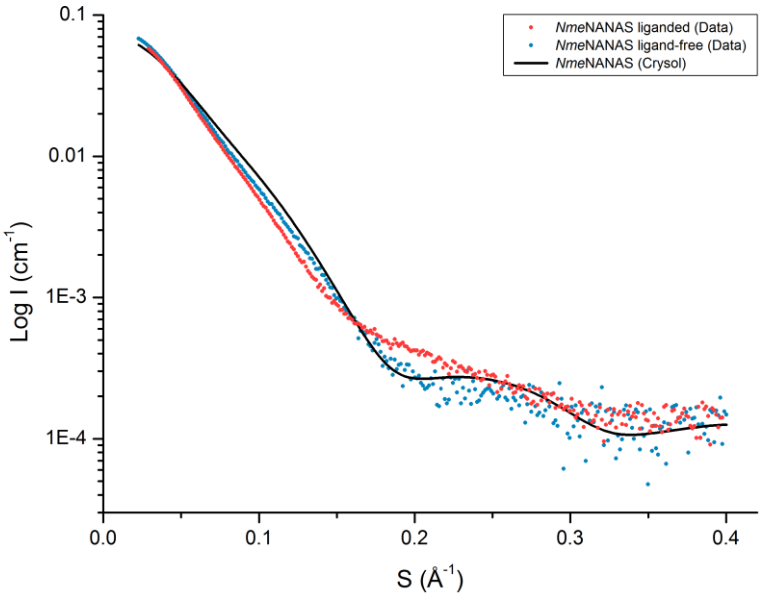


Figure 3.11 SAXS profiles of liganded and ligand-free wild-type *Nme*NANAS compared against theoretical CRYSOL scattering derived from crystallographic coordinates of *Nme*NANAS structure 1XUZ. In red is the experimental scattering profile for ligand-free *Nme*NANAS and in blue is the experimental scattering profile for *Nme*NANAS in the presence of 1 mM Mn^{2+} and 1 mM PEP. Theoretical scattering of *Nme*NANAS wild-type is depicted with a black line. SAXS data was collected using approximately 10 mg mL^{-1} protein in 30 mM triethanolamine-HCl buffer at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

	***Nme*NANAS**
I_0 (cm^{-1})	0.082 ± 0.001
R_g (Å)	33.0 ± 0.5
D_{max} (Å)	116
Porod volume ($Å^3$)	121,724
Estimated mass (kDa)	85
Theoretical mass of dimer (kDa)	77

Table 3.4 SAXS structural parameters determined for wild-type *Nme*NANAS in the presence of 1 mM Mn^{2+} and 1 mM PEP. SAXS data obtained at a wavelength of 1.0332 Å.

As shown in Figure 3.11, there is a discernible difference between the experimental scattering profiles determined for *Nme*NANAS wild-type in the presence and absence of Mn^{2+} and PEP. When compared to theoretical scattering generated from an *Nme*NANAS crystal structure (PDB code 1XUZ), it was observed that the experimental profile of liganded *Nme*NANAS wild-type protein fits significantly better than that for ligand-free enzyme (Chi values of 2.0 and 6.1 respectively). This result indicates that binding of Mn^{2+} and PEP is associated with a change in the average conformation of *Nme*NANAS. It can also be inferred that the solution structure of ligand-free *Nme*NANAS is different to the crystal structure, and that the crystal structure more closely resembles the Mn^{2+} and PEP bound solution conformation.

The structure parameters displayed in Table 3.4 correlate with a dimeric quaternary state for wild-type *Nme*NANAS. The molecular mass of *Nme*NANAS was calculated from SAXS data using SAXS MoW (*154*). The molecular mass of *Nme*NANAS estimated from SAXS was 85 kDa and is similar to that expected for dimeric enzyme (Table 3.4). The results from SAXS concur with the crystal structures of *Nme*NANAS (*112, 113*) and the results obtained from both analytical SEC (Section 3.5.3) as well as AUC (Section 3.5.4).

Unlike wild-type *Nme*NANAS, no obvious change was noted for the experimental profiles of *Nme*G272Term in the presence and absence of Mn^{2+} and PEP (Figure 3.12). In addition, both experimental scattering profiles for liganded and ligand-free *Nme*G272Term fit well to theoretical scattering generated from the truncated *Nme*NANAS crystal structure 1XUZ (Chi values of 0.52 and 0.66 respectively). These results indicate that the ability of PEP binding to induce a conformational change akin to what is observed for *Nme*NANAS wild-type is lost in *Nme*G272Term, and correlates with the observations made from DSF in Section 3.4.2, where *Nme*G272Term is not thermally stabilised by the addition of PEP.

The structure parameters obtained for *Nme*G272Term correlate with a dimeric quaternary state (Table 3.5). The molecular mass of *Nme*G272Term estimated from SAXS MoW (*154*) was 61 kDa and is similar to that expected for dimeric protein. The calculation of a dimeric molecular weight for *Nme*G272Term using SAXS data is likely due to the concentration of protein used for analysis. The solution structure of *Nme*G272Term was investigated by SAXS at approximately 10 mg/mL; a concentration at which the enzyme is most likely to be dimeric, given the results attained from analytical SEC and AUC (Sections 3.5.3 and 3.5.4).

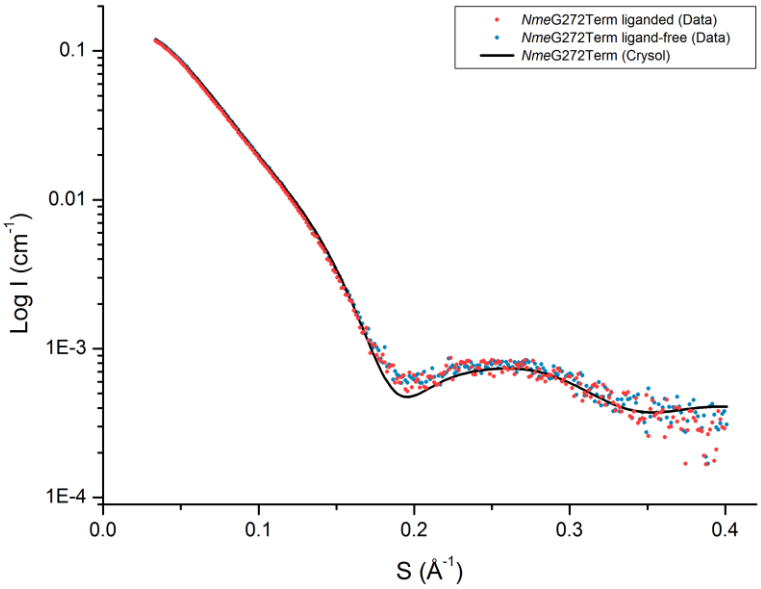


Figure 3.12 Small-angle X-ray scattering (SAXS) profiles of liganded and ligand-free *Nme*G272Term compared against theoretical CRYSOL scattering derived from modified crystallographic coordinates of *Nme*NANAS structure 1XUZ. In red is the experimental scattering profile for ligand-free *Nme*G272Term and in blue is the experimental scattering profile for *Nme*G272Term in the presence of 1 mM Mn^{2+} and 1 mM PEP. Theoretical scattering of *Nme*G272Term is depicted with a black line. SAXS data was collected using approximately 10 mg mL^{-1} protein in 30 mM triethanolamine-HCl buffer at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

	***Nme*G272Term**
I_0 (cm^{-1})	0.157 ± 0.001
R_g (Å)	26.7 ± 0.3
D_{max} (Å)	89
Porod volume ($Å^3$)	97,478
Estimated mass (kDa)	61
Theoretical mass of dimer (kDa)	60

Table 3.5 Small-angle X-ray scattering (SAXS) structural parameters determined for *Nme*G272Term in the presence of 1 mM Mn^{2+} and 1 mM PEP. SAXS data obtained at a wavelength of 1.0332 Å.

3.6 Molecular dynamics

In order to examine the difference in dynamic motions of wild-type *Nme*NANAS caused by PEP binding in the active site and the dynamic role of the AFPL domain, molecular dynamics (MD) simulations (270 ns) were conducted for *Nme*NANAS with and without PEP present. MD simulations as well as the following analyses were carried out by Dr Wanting Jiao.

The simulations of *Nme*NANAS with and without PEP both equilibrated after 100 ns. The average structures from both simulations were generated using MD trajectories from the equilibrated time period (100 – 270 ns). The MD simulations for *Nme*NANAS with or without PEP bound both started with the "closed" conformation of the active site, based on the crystal structure that contains both PEP and rManNAc in the active site (PDB 1XUZ). During the MD simulation for ligand-free *Nme*NANAS, the active site loops that contain ManNAc binding residues (Gln55 to Asn89) and the AFPL domain linker (residues Ala270 to Ile279) in both monomers were found to open up, (Figure 3.13A and Figure 3.13B). As a result, the critical Arg314 residue in both monomers has shifted outwards, by 4.2 Å and 2.2 Å relative to its starting positions in chain A and chain B respectively.

In the MD simulation with PEP bound in the active site, the two monomers exhibit slightly different behaviours. In chain A (Figure 3.13C), the linker region showed similar outwards motion as in the ligand-free system, and results in an outward shift of 2.2 Å of residue Arg314 relative to the starting position. In chain B (Figure 3.13D), interestingly, the main AFPL domain (residues 291 to 349) moved towards the main barrel, in contrast to what is observed in the calculations for chain A and the ligand-free system. This movement of the AFPL domain did not result in Arg314 being pulled away from the active site as for chain A; the distance shifted by Arg314 was only 0.6 Å relative to the starting position in chain B. Therefore, it appears that PEP binding restricts the movement of the AFPL domain and its linker, so that the position of the catalytically critical Arg314 residue is held closer to the active site (0.6 Å and 2.2 Å in PEP bound *Nme*NANAS compared to 2.2 Å and 4.2 Å in the ligand-free *Nme*NANAS relative to their starting positions).

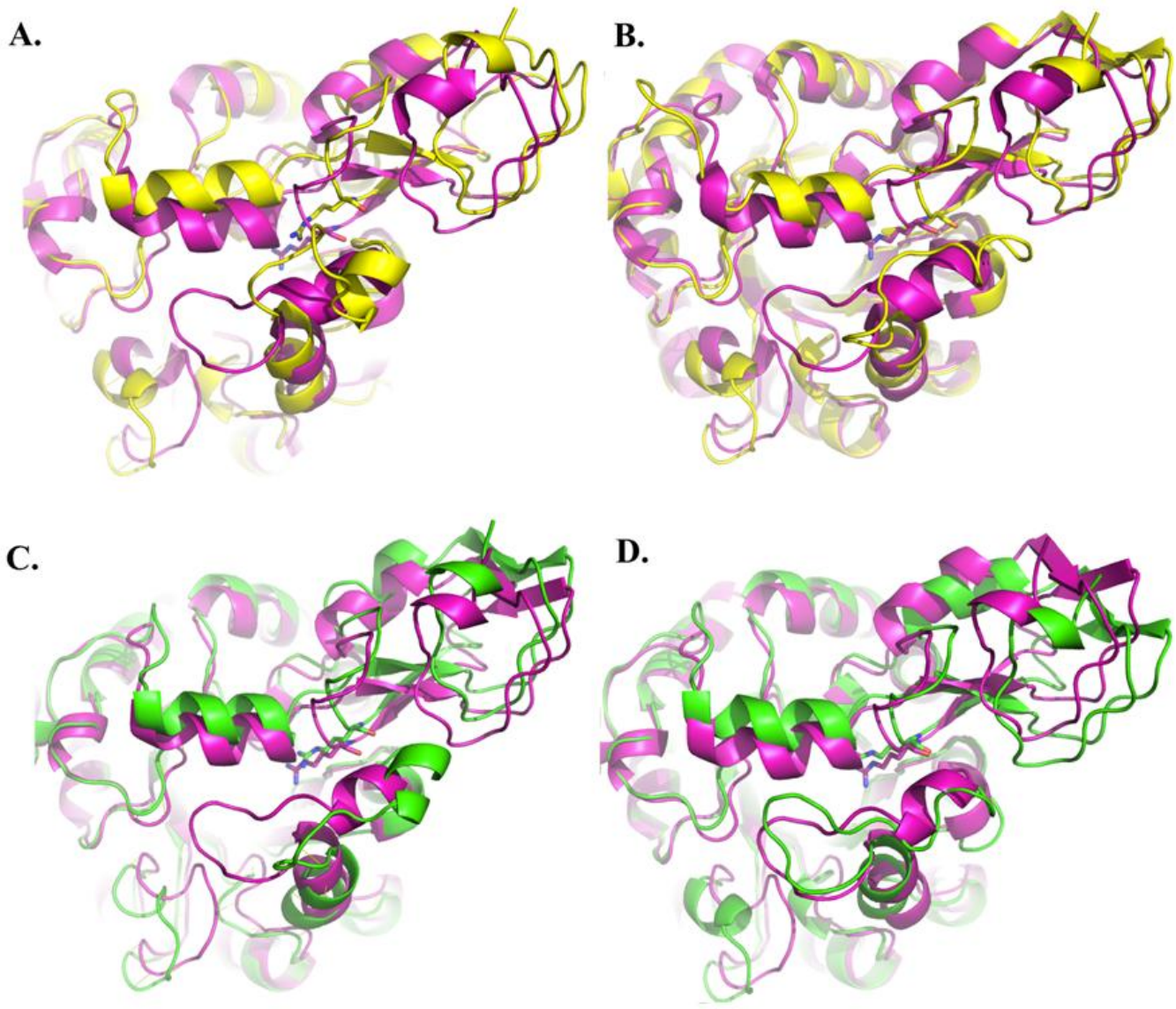


Figure 3.13 Conformations of the ManNAc loop and the AFPL domain superimposed between A) chain A of average structure from MD simulation of ligand-free *Nme*NANAS (yellow) and crystal structure (purple); B) chain B of average structure from MD simulation of ligand-free *Nme*NANAS (yellow) and crystal structure (purple); C) chain A of average structure from MD simulation for PEP bound *Nme*NANAS (green) and crystal structure (purple); D) chain B of average structure from MD simulation for PEP bound *Nme*NANAS (green) and crystal structure (purple). Residue Arg314 is displayed with sticks representation. Figure created by Dr Wanting Jiao.

The chain average RMSF values for PEP bound and ligand-free *Nme*NANAS were calculated in order to investigate any change in flexibility of the enzyme associated with PEP binding (Figure 3.14C). The difference in RMSF values indicate that the flexibility of the ManNAc binding, β2-α2 loop and AFPL domain regions are reduced significantly upon PEP binding (Figure 3.14D).

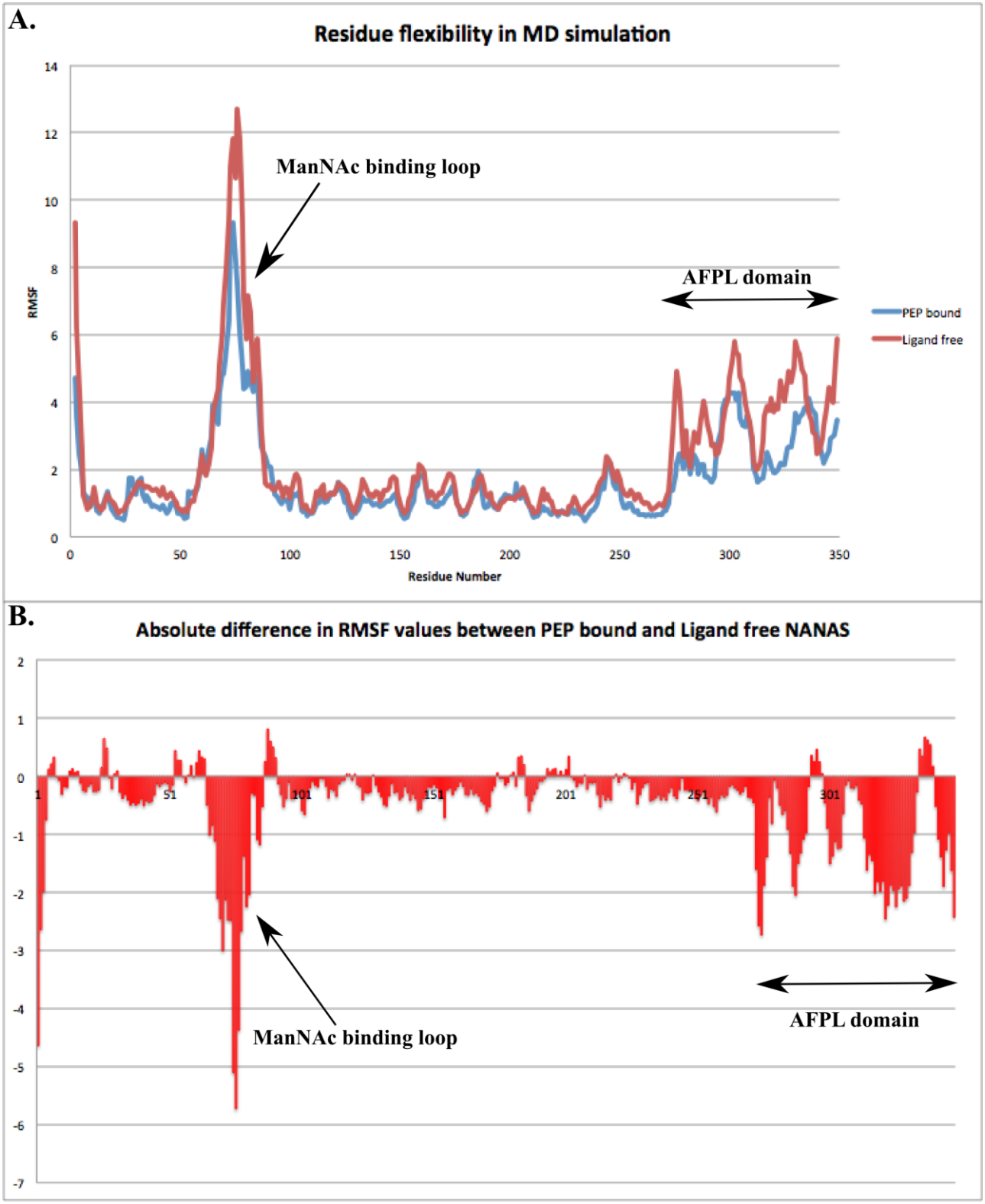


Figure 3.14 (A) Chain average residue RMSF values during MD simulations for PEP bound and ligand-free NANAS. (B) Difference in chain average RMSF values between PEP bound and ligand-free *Nme*NANAS. Figure created by Dr Wanting Jiao.

The interactions between the AFPL domain (residues Gly272 to Glu349) and the main barrel excluding the ManNAc binding loop (Gln55 to Asn89) during the equilibrated time period in MD simulations were examined. In the PEP bound system, analyses identified four interactions between the main barrel and the AFPL domain, Arg114-Asp296, Arg194-Glu282, Asn184-Thr285 and Arg314-Glu134, with high occupancy (Figure 3.15, table of data from hydrogen bond analysis is in Table 3.6). Interestingly, these four interactions are distributed across the structure of the AFPL domain, so that the domain is well anchored to the main barrel. The interaction between Arg194 and Asp282 is particularly interesting because Arg194 belongs to the PEP binding loop, and any changes in this loop caused by PEP binding may be communicated through to the AFPL domain through this key interaction. There is a network of hydrogen bonds formed around this region, involving the backbone of Asn184 and side-chains of Asp282 and Thr285 (Figure 3.15). In the crystal structure *Nme*NANAS 1XUZ, the side-chain of Asn184 forms a hydrogen bond with the phosphate group of PEP (Figure 3.1B).

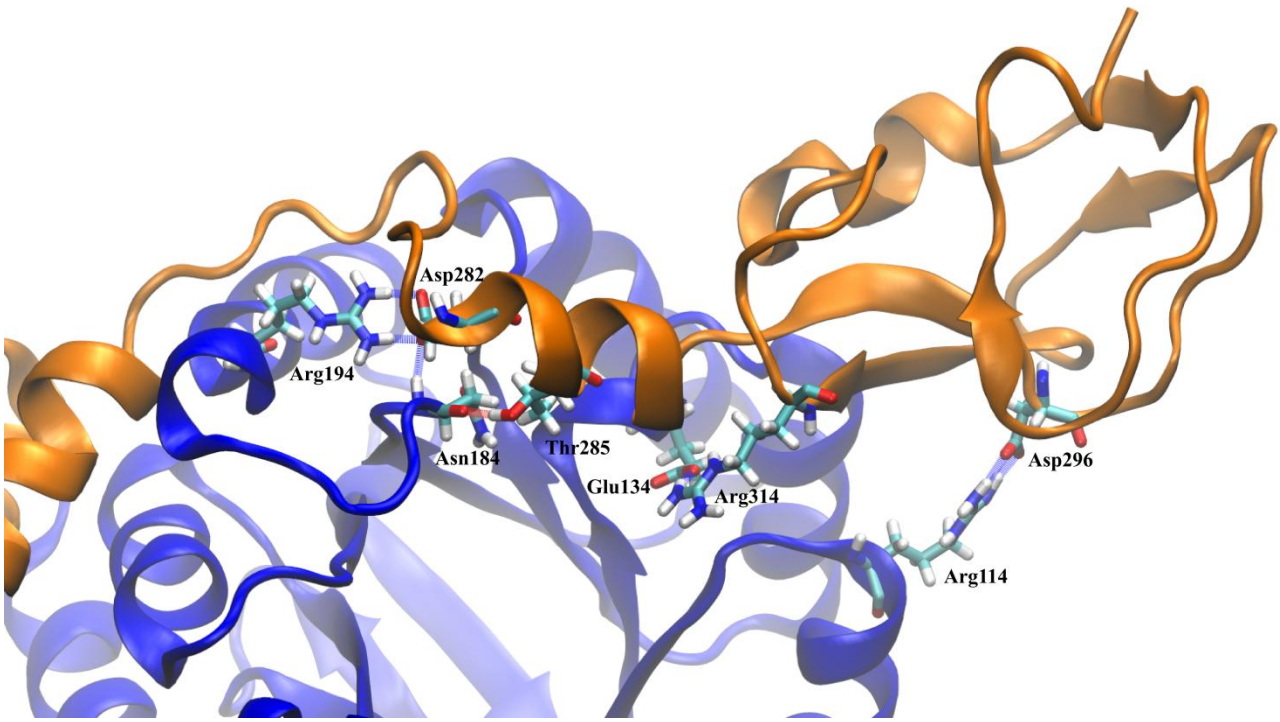


Figure 3.15 Key hydrogen bonding interactions between main barrel and the AFPL domain observed from MD simulations. Figure created by Dr Wanting Jiao.

Chain B barrel & Chain A AFPL domain			Chain A barrel & Chain B AFPL domain		
donor	**acceptor**	**occupancy**	**donor**	**acceptor**	**occupancy**
Arg314-Side	Glu134-Side	71.07%	Arg314-Side	Glu134-Side	52.78%
Arg114-Side	Asp296-Side	68.93%	Arg114-Side	Asp296-Side	88.04%
Asn184-Main	Glu282-Side	40.96%	Asn184-Main	Glu282-Side	38.70%
Arg194-Side	Glu282-Side	31.96%	Arg194-Side	Glu282-Side	85.26%
Thr285-Side	Asn184-Main	31.26%	Thr285-Side	Asn184-Main	44.37%
Asn184-Side	Glu282-Side	12.85%	Asn184-Side	Glu282-Side	3.52%

Table 3.6 Key hydrogen bonds and their occupancy between the main barrel and the AFPL domain from both monomers. Table created by Dr Wanting Jiao.

The same hydrogen bond analysis was conducted for the ligand-free *Nme*NANAS system, which identified similar hydrogen bonds with high occupancy to those found in the PEP bound system (Table 3.7). The most significant difference in occupancy occurs for the interaction between Arg314 and Glu134, the occupancy of which increases by more than 40% upon PEP binding. This change in occupancy suggests that PEP binding in the active site indeed affects how Arg314 interact with the surrounding residues and how this residue is held in the active site.

	PEP bound			Ligand-free		
	Donor	**Acceptor**	**Occupancy**	**Donor**	**Acceptor**	**Occupancy**
Chain A barrel	Arg114-Side	Asp296-Side	88.04%	Arg114-Side	Asp296-Side	83.81%
& Chain B	Arg194-Side	Glu282-Side	85.26%	Arg194-Side	Glu282-Side	85.48%
AFPL domain	Arg314-Side	Glu134-Side	52.78%	Arg314-Side	Glu134-Side	0.37%
Chain B barrel	Arg114-Side	Asp296-Side	68.93%	Arg114-Side	Asp296-Side	76.22%
& Chain A	Arg194-Side	Glu282-Side	31.96%	Arg194-Side	Glu282-Side	1.33%
AFPL domain	Arg314-Side	Glu134-Side	71.07%	Arg314-Side	Glu134-Side	28.63%

Table 3.7 Hydrogen bonds occupancies in ligand-free and PEP bound *Nme*NANAS during MD simulations. Table created by Dr Wanting Jiao.

3.7 Cloning, expression and purification of *Nme*NANAS hydrogen bond network variants

3.7.1 Site-directed Mutagenesis

In order to investigate the role of the AFPL domain, several point mutations of *Nme*NANAS were generated using residues identified from the MD simulations of wild-type apoenzyme and enzyme bound to Mn^{2+} and PEP (Section 3.6). Residues Thr285, Glu282 and Glu134 were each substituted to Ala to completely remove side-chain hydrogen bonding capability. Residue Thr282 was also mutated to Phe to remove hydrogen bonding functionality but also preserve a similar steric presence. The final variant was a double mutation of Thr285 and Glu282 to alanine to investigate if their combined loss of functionality would affect catalysis.

The *Nme*NANAS wild-type pDEST-17 vector described in Chapter 2 was used as a template to generate the *Nme*NANAS hydrogen bond network variants. The methods used for PCR amplification of the variant genes are outlined in Section 8.3.2, and primer sequences used for site-directed mutagenesis are displayed in Table 8.7.

3.7.2 Expression and purification

The pDEST-17 vectors bearing the variant genes were transformed into Chaperone 3 cells to optimise the yield of soluble protein. *Nme*E134A, *Nme*E282A, *Nme*T285A, *Nme*T285F, and *Nme*E282A/T285A were each expressed and purified using a similar method to that outlined for *Nme*NANAS wild-type in Section 8.4. Purified enzymes were concentrated, flash-frozen and stored at -80 °C. Samples of purified enzyme were analysed by SDS PAGE (Figure 3.16).

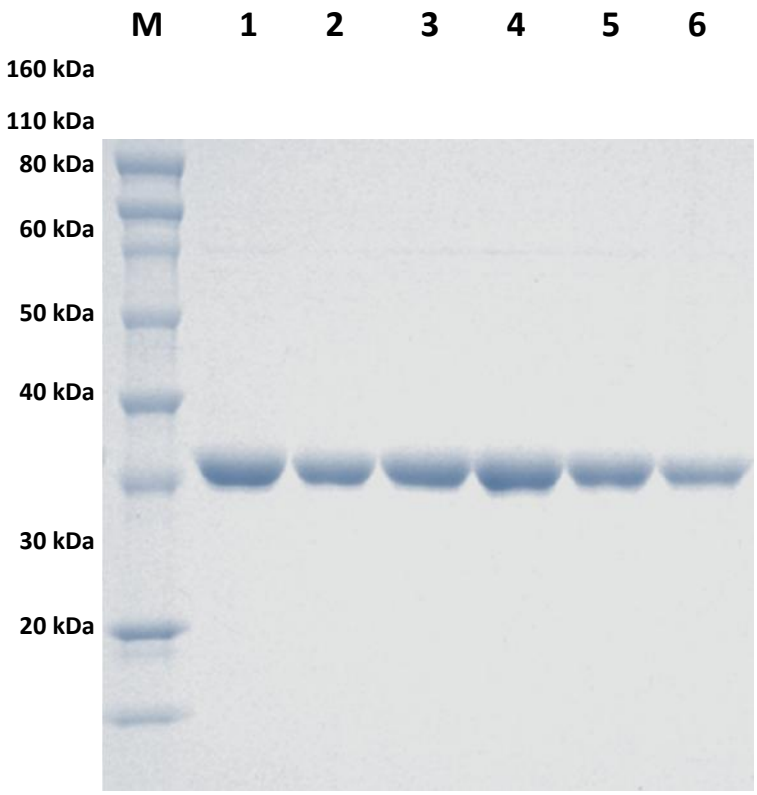


Figure 3.16 SDS-PAGE gel of purified *Nme*NANAS wild-type and hydrogen bond network variants. Lane M contains marker with molecular weights defined. Lane 1 contains purified *Nme*NANAS wild-type, Lane 2 *Nme*E134A, Lane 3 *Nme*E282A, Lane 4 *Nme*T285A, Lane 5 *Nme*T285F and Lane 6 *Nme*E282A/T285A.

3.8 Physical characterisation of *Nme*NANAS hydrogen bond network variants

3.8.1 Circular dichroism

CD was used to assess and compare the secondary structure of the hydrogen bond network variants with wild-type *Nme*NANAS (Figure 3.17).

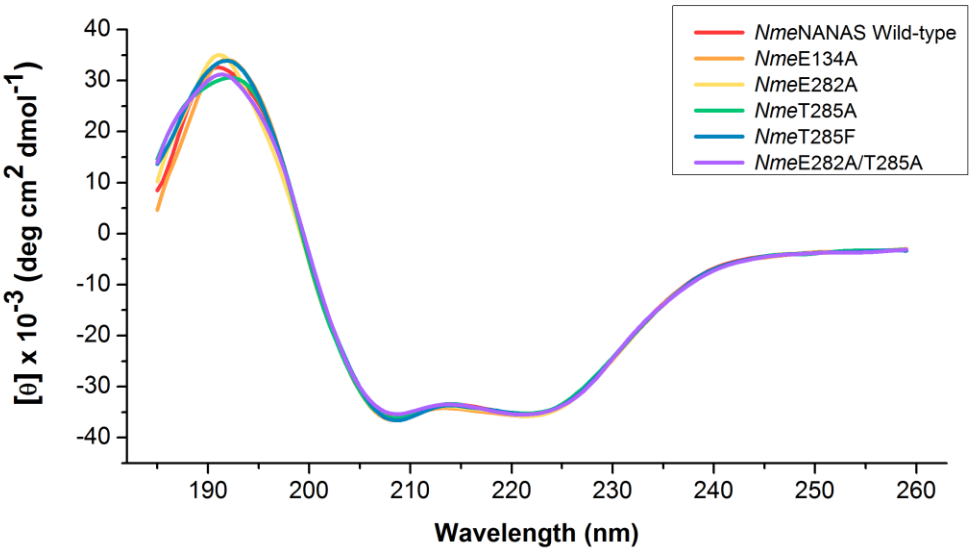


Figure 3.17 CD spectra for *Nme*NANAS wild-type (red), *Nme*E134A (orange), *Nme*E282A (yellow), *Nme*T285A (green), *Nme*T285F (blue) and *Nme*E282A/T285A (purple).

Displayed in Figure 3.17 the CD profiles for each of the *Nme*NANAS hydrogen bond network variants are almost identical to the profile for *Nme*NANAS wild-type. Furthermore, the secondary structure ratios determined from the variant CD data using the K2D3 server (*142*), were nearly identical to the values obtained for wild-type *Nme*NANAS. The results from CD thereby indicate that the secondary structure of the variants is unlikely to have been significantly affected by any of the mutations made.

3.8.2 Mass spectrometry

The masses of each of the *Nme*NANAS hydrogen bond network variants were analysed using mass spectrometry (Table 3.8). Each of the experimental masses obtained were almost identical to those calculated from the respective peptide sequences using ProtParam (*141*).

3.8.3 Thermal stability

DSF was used to assess the melting temperatures of the *Nme*NANAS hydrogen bond network variants. The method used for DSF analysis is outlined in Section 8.5.3. Displayed in Table 3.8, the melting points for apoenzyme *Nme*E134A, *Nme*E282A, *Nme*T285A, *Nme*T285F, and *Nme*E282A/T285A were approximately 43 °C, similar to that determined for wild-type *Nme*NANAS (Section 3.4.2). The results from DSF indicate that the thermal stability of the variant proteins have not been significantly altered through mutagenesis.

	***Nme*NANAS wild-type**	***Nme*E134A**	***Nme*E282A**	***Nme*T285A**	***Nme*T285F**	***Nme*E282A/ T285A**
Experimental mass (Da)	41,977	41,919	41919	41,947	42023	41,889
Theoretical mass (Da)	41,977	41,918	41,917	41,947	42,023	41,888
Melting Temperature (°C)	42.7 ± 0.2	43.0 ± 0.5	42.9 ± 0.2	43.4 ± 0.4	43.3 ± 0.2	42.2 ± 0.2

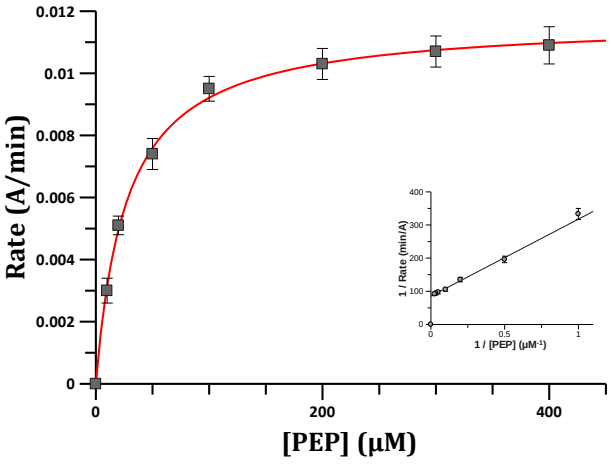
Table 3.8 Mass spectrometry measurements and melting temperatures of His_6-tagged *Nme*E134A, *Nme*-E282A, *Nme*T285A, *Nme*T285F and *Nme*E282A/T285A.

3.9 Kinetic characterisation of *Nme*NANAS hydrogen bond network variants

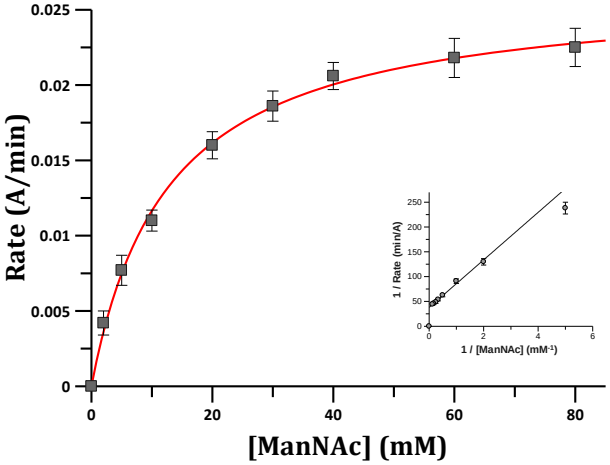
3.9.1 Michaelis-Menten kinetics

To determine the kinetic parameters of the *Nme*NANAS hydrogen bond network variants, the consumption of PEP was monitored using the assay described in Section 8.5.1. Cuvettes containing 50 mM BTP (pH 8), 1 mM $MnCl_2$, PEP (varied) and ManNAc (varied) were incubated for 10 min at 25 °C. The enzymatic reaction was initiated by the addition of purified enzyme. Michaelis-Menten steady-state approximations were utilised to determine kinetic parameters (Figure 3.18). Kinetic parameters were compared to the values ascertained for *Nme*NANAS wild-type (Table 3.9).

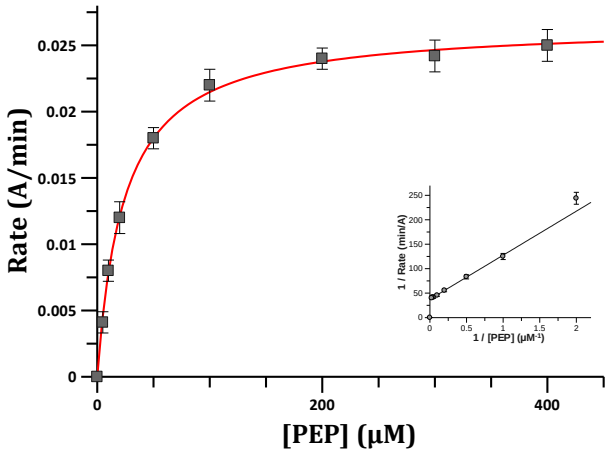
A


*Nme*E134A: K_m (PEP)
Rate (A/min)
[PEP] (µM)
1 / Rate (min/A)
1 / [PEP] (µM-1)

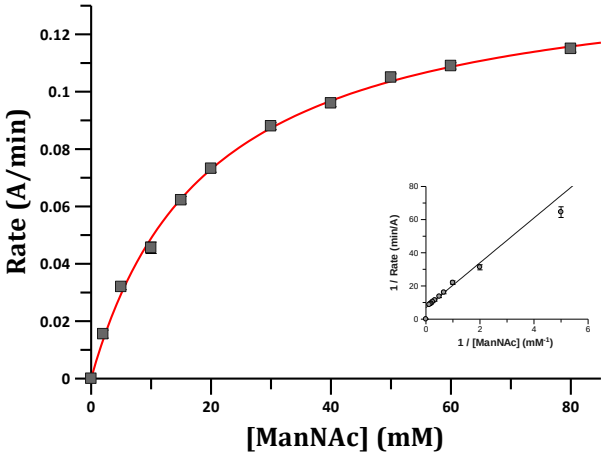
B


*Nme*E134A: K_m (ManNAc)
Rate (A/min)
[ManNAc] (mM)
1 / Rate (min/A)
1 / [ManNAc] (mM-1)

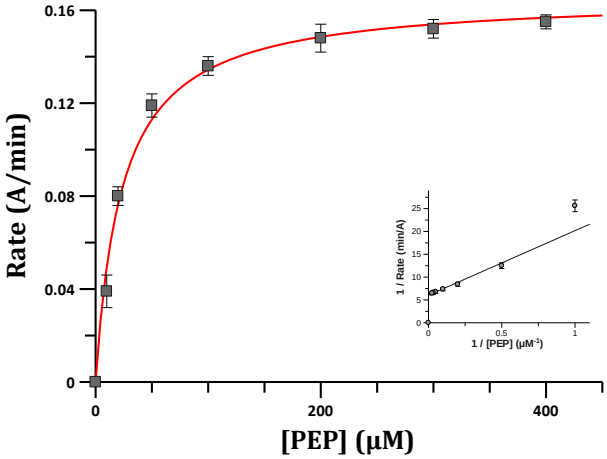
C


*Nme*E282A: K_m (PEP)
Rate (A/min)
[PEP] (µM)
1 / Rate (min/A)
1 / [PEP] (µM-1)

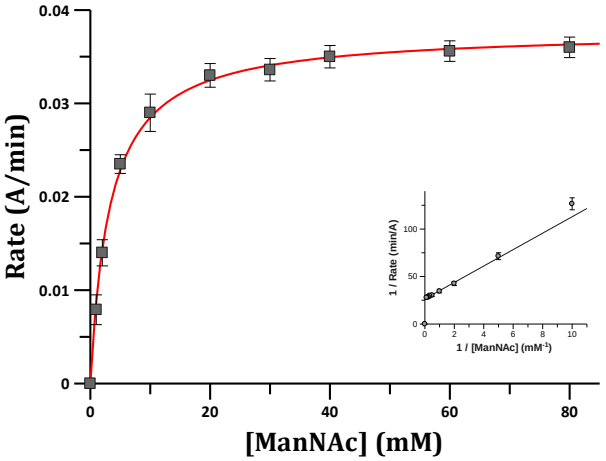
D


*Nme*E282A: K_m (ManNAc)
Rate (A/min)
[ManNAc] (mM)
1 / Rate (min/A)
1 / [ManNAc] (mM-1)

E


*Nme*T285A: K_m (PEP)
Rate (A/min)
[PEP] (µM)
1 / Rate (min/A)
1 / [PEP] (µM-1)

F


*Nme*T285A: K_m (ManNAc)
Rate (A/min)
[ManNAc] (mM)
1 / Rate (min/A)
1 / [ManNAc] (mM-1)

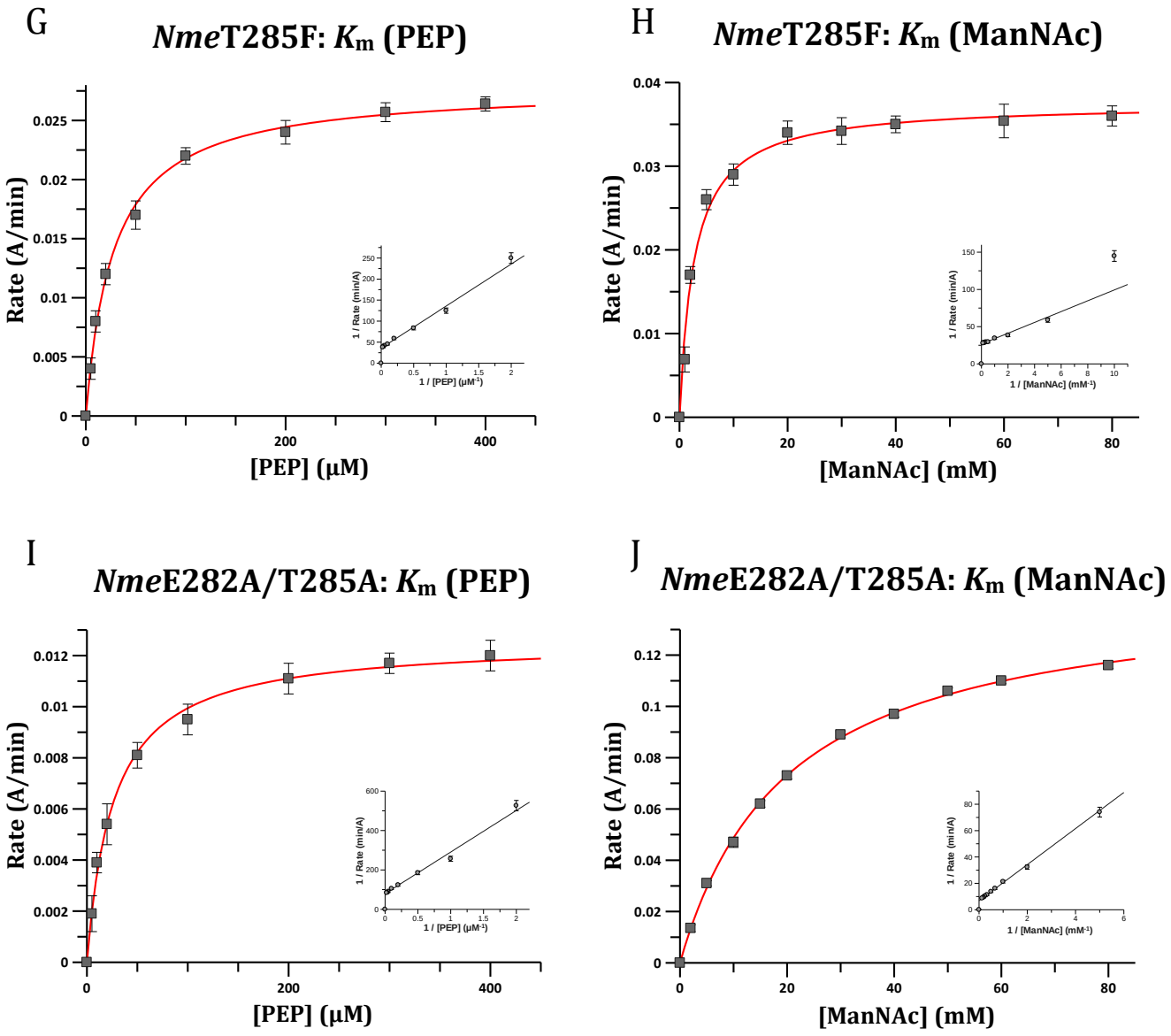


Figure 3.18 Michaelis-Menten curves for *Nme*NANAS hydrogen bond network variants with ManNAc and PEP. (A) *Nme*E134A PEP K_m curve, (B) *Nme*E134A ManNAc K_m curve, (C) *Nme*E282A PEP K_m curve, (D) *Nme*E282A ManNAc K_m curve, (E) *Nme*T285A PEP K_m curve, (F) *Nme*T285A ManNAc K_m curve, (G) *Nme*T285F PEP K_m curve, (H) *Nme*T285F ManNAc K_m curve, (I) *Nme*E282A/T285A PEP K_m curve and (J) *Nme*E282A/T285A ManNAc K_m curve. Data points are shown in black and a line of best fit in red. A double reciprocal plot is displayed as an inset for each graph. Assays were carried out in 50 mM BTP at pH 7.5 with 1 mM $MnCl_2$, PEP (varied) and ManNAc (varied) at 25 °C. Reaction was initiated by the addition of 5 µL of purified enzyme at a stock concentration of 2.5 – 3.5 mg/mL. Michaelis-Menten steady-state approximations were utilised to determine kinetic parameters with nonlinear fitting in GraFit (Erathicus Software). To determine apparent K_m (PEP), ManNAc concentration was held at 30 mM, and the concentration of PEP varied. To determine apparent K_m (ManNAc), PEP was kept constant at 1 mM, and ManNAc concentration was varied.

	Wild-type	G272Term	T285A	T285F	E282A	E282A/T285A	E134A
k_{cat} (s^{-1})	3.1 ± 0.1	-	2.8 ± 0.1	3.0 ± 0.1	0.082 ± 0.003	0.085 ± 0.002	0.3 ± 0.01
K_m PEP (μM)	28 ± 3	-	24 ± 2	27 ± 2	24 ± 1	26 ± 2	27 ± 2
K_m ManNAc (mM)	2.9 ± 0.2	-	3.1 ± 0.3	2.8 ± 0.3	20.1 ± 1.6	21.0 ± 1.9	13 ± 1
k_{cat}/K_m ManNAc (s^{-1}/M)	1068 ± 108	-	900 ± 120	1071 ± 150	4.7 ± 0.5	4.4 ± 0.5	23 ± 3

Table 3.9 Kinetic parameters for wild-type *Nme*NANAS, *Nme*G272Term, *Nme*E134A, *Nme*E282A, *Nme*-T285A, *Nme*T285F and *Nme*E282A/T285A. Values with - indicates k_{cat} of enzyme is less than 0.004 s^{-1}.

Depicted in Table 3.9, each of the variants retain similar K_m (PEP) values to that observed in wild-type *Nme*NANAS, indicating that the PEP binding site has not been affected by mutation. Variants *Nme*T285A and *Nme*T285F demonstrated very similar kinetic parameters to the wild-type enzyme (Table 3.9). This indicates that the mutation of Thr285 has no effect upon the catalytic function of *Nme*NANAS.

In contrast, both *Nme*E282A and the double variant *Nme*E282A/T285A exhibited an approximately 230-fold reduction in k_{cat}/K_m (ManNAc) compared to wild-type enzyme. This results from an approximately 35-fold reduction in k_{cat} and 7-fold increase in the K_m (ManNAc) for both *Nme*E282A and *Nme*E282A/T285A. The similar kinetic values for *Nme*E282A and *Nme*E282A/T285A correlate with the observation that Thr285 is not essential to the catalytic function of *Nme*NANAS as no additional kinetic penalty is afforded through the double mutation when compared with *Nme*E282A (Table 3.9).

Variant *Nme*E134A has a somewhat less drastic affect on catalysis when compared with *Nme*E282A. *Nme*E134A has an approximately 45-fold reduction in k_{cat}/K_m (ManNAc) when compared to wild-type enzyme, arising from an approximately 10-fold reduction in k_{cat} and 4-fold increase in K_m (ManNAc).

Glu282, Glu134 and Thr285 are each fully conserved in NANAS enzymes; however, T285 is not conserved in other members of the SAS family such as LegSs and PseSs. It is thereby likely that Glu282 and Glu134 are involved in the ManNAc site pre-organisation and hydrogen bond relay whereas Thr285 is not.

3.10 Summary of results

An AFPL domain truncation of *Nme*NANAS was successfully amplified and cloned into a His-tag expression vector, and subsequently expressed and purified by IMAC and SEC. *Nme*G272Term was found to be soluble but inactive using the continuous assay previously described for wild-type *Nme*NANAS in Chapter 2.

Binding studies using ITC identified that the Mn^{2+} and PEP binding sites of *Nme*G272Term were intact however the protein was unable to bind rManNAc. The loss of rManNAc binding indicates that the AFPL domain plays an important role in this process for wild-type *Nme*NANAS. In addition, the thermal stabilities of *Nme*NANAS wild-type and *Nme*G272Term were assessed in the presence of various substrate combinations. The results from DSF revealed a significant thermal stabilisation of wild-type enzyme in the presence of PEP which was absent in *Nme*G272Term.

Unlike wild-type *Nme*NANAS which is a tight dimer, *Nme*G272Term was proven to have a concentration dependent, monomer-dimer equilibrium using analytical SEC and AUC. The results from analytical SEC and AUC thus strongly imply that the AFPL domain plays an important role in stabilising the quaternary structure of wild-type *Nme*NANAS.

Several attempts to crystallise *Nme*G272Term were made however no diffraction quality crystals were obtained. The solution structures of wild-type *Nme*NANAS and *Nme*G272Term were assessed in the absence and presence of Mn^{2+} and PEP using SAXS. Wild-type *Nme*NANAS was observed to undergo conformational changes in the presence of Mn^{2+} and PEP whereas *Nme*G272Term was not. These results strongly imply that the AFPL domain is repositioned upon binding of PEP and correlates with the premise that PEP liganding pre-organises the ManNAc binding site (Chapter 2).

Molecular dynamics simulations of wild-type *Nme*NANAS with and without PEP were able to identify a hydrogen bonding relay, linking the AFPL domain and catalytic domains. Key residues involved in the hydrogen bond network were mutated and variant enzymes were generated and kinetically analysed. The conserved residues Glu134 and Glu282 were identified to play a key role in the function of the hydrogen bonding relay, as loss of their functionality was observed to significantly detriment catalytic function.

The results from this chapter identify that the AFPL domain of *Nme*NANAS is bifunctional, playing both catalytic and structural roles. The AFPL domain is linked to the catalytic domain via a hydrogen bonding network and is crucial to the mechanism by which the ManNAc binding site is pre-organised by PEP. The AFPL domain is also involved in maintaining a stable quaternary state in wild-type *Nme*NANAS.

3.10 Discussion

The work in Chapter 3 details an investigation into the role of the AFPL domain of *Nme*NANAS This was achieved through the generation of an AFPL domain truncation *Nme*NANAS variant. *Nme*G272Term was found to be soluble. ITC binding studies with *Nme*G272Term were able to determine that the truncated variant could still bind Mn^{2+} and PEP, but that the ability to bind rManNAc had been lost. The AFPL domain is therefore critical to ManNAc binding in *Nme*NANAS.

DSF analysis revealed a significant thermal stabilisation of wild-type *Nme*NANAS in the presence of PEP. Additionally, SAXS data revealed that wild-type *Nme*NANAS undergoes a conformational change when PEP is bound. Neither thermal stabilisation nor conformational change were observed with *Nme*G272Term in the presence of PEP, implying that the mechanism by which PEP pre-organised the ManNAc binding site in wild-type *Nme*NANAS had been lost in *Nme*G272Term. A hydrogen bonding relay linking the AFPL domain and catalytic domain was identified through MD simulations of wild-type *Nme*NANAS. Residues Glu134 and Glu282 were identified though kinetic characterisation of variant enzymes, to play a key role in the function of the aforementioned hydrogen bonding relay.

Analytical SEC and AUC revealed that the AFPL domain plays a vital structural role, as loss of the AFPL domain in *Nme*G272Term resulted in a schizophrenic quaternary structure. The results described in Chapter 3 thus identified a bifunctional role of the *Nme*NANAS AFPL domain, which is involved in both the pre-organisation of the ManNAc binding site via a hydrogen bonding relay, as well as stabilisation of the enzyme's dimeric state.

Chapter 4

Evolutionary relationships within the bacterial sialic acid synthase family

4.1 Background

4.1.1 Bacterial sialic acid synthases

NANAS is the most extensively researched member of the SAS family (*5*). The crystal structure of NANAS from *N. meningitidis* is the only wild-type SAS structure currently elucidated (*112, 113*). PseS and LegS are also members of the SAS family (*64, 68, 89, 135, 138*). Shown in Figure 4.1, LegS catalyses an aldol-like condensation reaction between PEP and 2,4-diacetamido-2,4,6-trideoxymannose to form Leg and PseS reacts PEP with 2,4-diacetamido-2,4,6-trideoxyaltrose to form Pse.

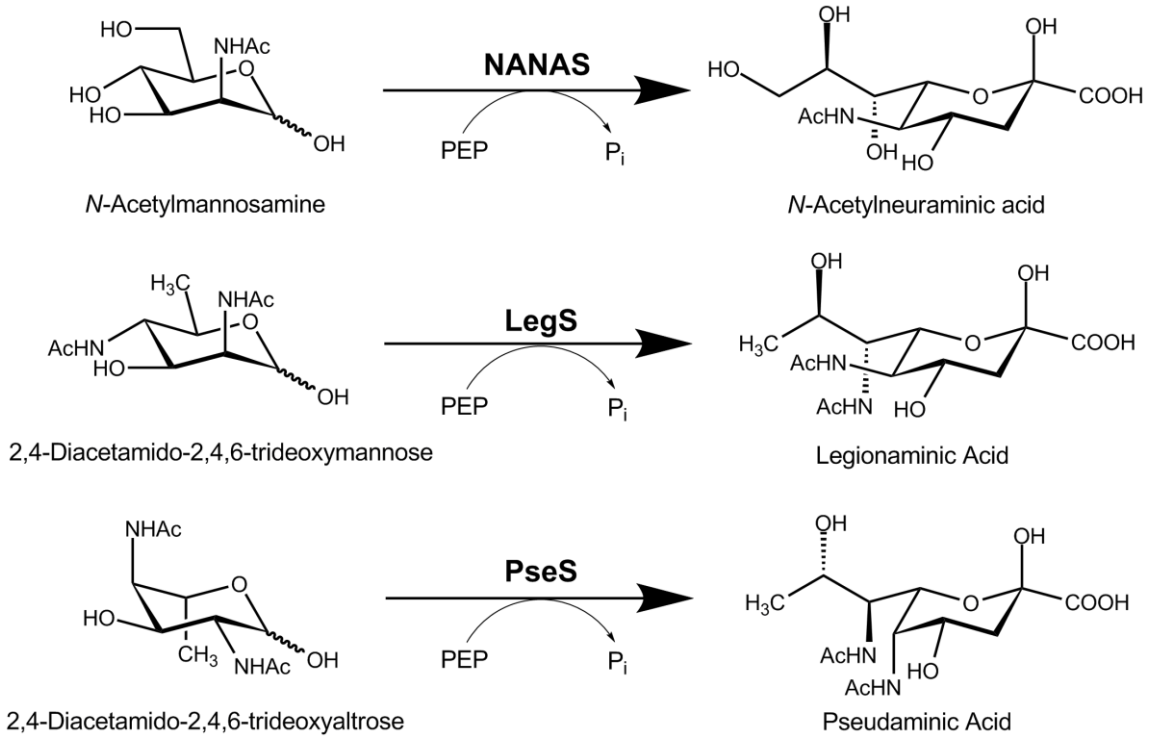


Figure 4.1 Enzymatic reactions catalysed by *N*-acetylneuraminic acid synthase (NANAS), legionaminic acid synthase (LegS) and pseudaminic acid synthase (PseS).

Leg has the same absolute configuration at each of its stereocentres as *N*-acetylneuraminic acid (NANA) (Figure 4.1). The structure of Leg differs from NANA at position C-7 where Leg has a second *N*-acetyl group and C-9 where NANA has a hydroxyl group but Leg does not. Pse is a diastereoisomer of Leg with inverted stereochemistry at positions C-5, C-7 and C-8. Both PseS and LegS have been studied previously to a much lesser extent than NANAS and thus very little is currently known of the evolutionary relationships between these enzymes.

4.1.2 Goals of this research

This research aims to further develop the understanding of the structural and functional relationships within the family of SAS enzymes. The first step in this investigation required the analysis of available bacterial SAS sequences in an effort to uncover evolutionary relationships. Three different SAS genes are found in *C. jejuni* (*64*). The enzyme encoded by *Cj1141* has been characterised as a NANAS (*90*), *Cj1327* as a LegS (*68*) and *Cj1317* as a PseS (*89*). The existence of each of these genes in the same organism provides a useful model for investigating the evolutionary relationships between these members of the SAS family. The SAS enzyme from *C. tepidum* (a green sulphur bacterium first isolated from acidic sulphide hot springs in New Zealand (*160*)) was selected to investigate a member of the SAS family which naturally lacked and AFPL domain.

4. 2 Sequence and Phylogenetic Analysis of SASs

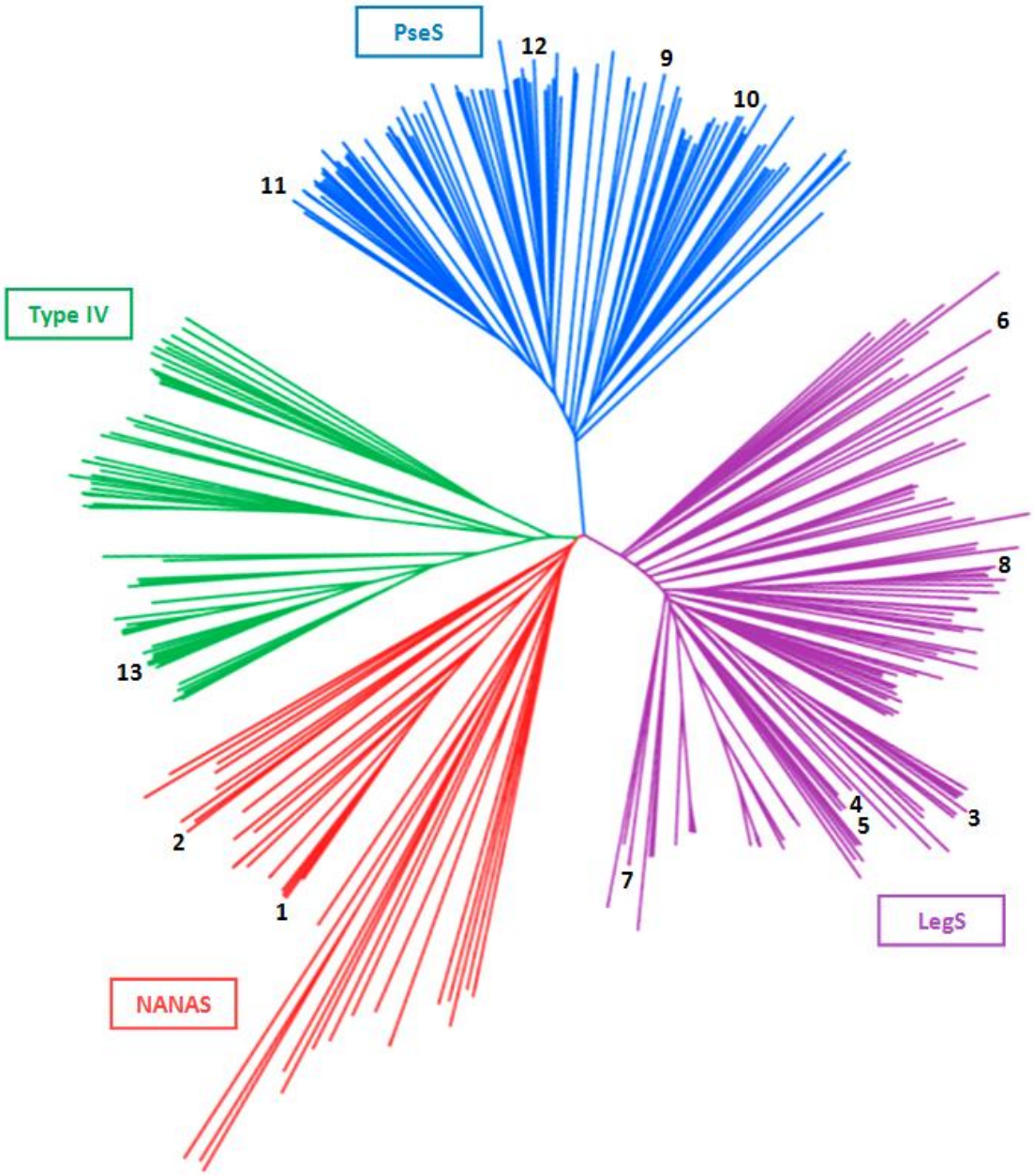
4.2.1 Sialic acid synthase phylogenetics

In order to determine the evolutionary relationships within the sialic synthase (SAS) family, an extensive sequence alignment of 410 known and potential SASs genes was constructed using ClustalW2 (*155, 156*). The methods used for sequence selection are outlined in Section 8.8.2. A phylogenetic tree was generated from the output of the ClustalW2 alignment using FigTree (*157*) (Figure 4.2). The sequence identity of SASs genes used in the alignment varied from 98.7 % to 21.1 %.

The SAS phylogenetic tree (Figure 4.2) identifies four major clades within the SAS family. These clades have been assigned in this thesis as NANAS, LegS, PseS and a fourth grouping of enzymes with unkown activity labelled as Type IV. The cluster termed as NANAS enzymes contains the sequences for *Nme*NANAS and *Cje*NANAS which have both been previously reported to have NANAS activity (*90, 112*).

The largest clade is labelled as the LegS subfamily. Within this cluster are both *Lpn*LegS and *Cje*LegS which have both been previously reported to have LegS activity (*68, 71, 135*). Futhermore this grouping contains the SAS from *Photobacterium profundum* strain 3TCK which was identified to produce Leg but not NANA or Pse through high-performace liquid chromatography (HPLC) and mass spectroscopy (MS) experiments (*134*).

Interestingly, the LegS clade also contains the SAS enzymes from both *E. coli* and *Streptococcus agalacticae*. Both of these enzymes have been shown to have NANAS activity (*95, 105, 158*) yet they are grouped within the LegS subfamily and not the NANAS subfamily. The SAS from *S. agalacticae* has 44.5 % and 52.5% sequence identity with *Lpn*LegS and *Cje*LegS respectively and only 33.1% sequence identity with *Nme*NANAS. Similarly, the SAS synthase from *E. coli* has 41.5 % and 55.2 % sequence identity with *Lpn*LegS and *Cje*LegS respectively and only 33.1 % sequence identity with *Nme*NANAS. Unfortunately, neither *Eco*SAS nor *Sag*SAS have been tested for LegS activity.



	Organism	SAS activity
1	*Neisseria meningitidis*	NANAS (*5, 103, 112*)
2	*Campylobacter jejuni* (Cj1141)	NANAS (*90*)
3	*Escherichia coli*	NANAS (*95, 158*)
4	*Streptococcus agalacticae*	NANAS (*105*)
5	*Streptococcus suis*	NANAS
6	*Legionella pneumophila*	LegS (*68*)
7	*Campylobacter jejuni* (Cj1327)	LegS (*135*)
8	*Photobacterium profundum* strain 3TCK	LegS (*134*)
9	*Campylobacter jejuni* (Cj1317)	PseS (*89, 138*)
10	*Methanobrevibacter smithii*	PseS (*134*)
11	*Photobacterium profundum* strain SS9	PseS (*134*)
12	*Helicobacter pylori*	PseS (*63*)
13	*Chlorobium tepidum*	-

Figure 4.2 Phylogenetic tree of bacterial SAS. In red are NANASs, purple are LegSs, blue are PseSs and green are Type IV SASs. Numbers depict SASs which have been identified in the literature. Organism name and known activity is tabulated above. Activities marked with a - indicate that the reaction is currently unknown.

The grouping of enzymes observed as both NANAS and LegS within the same clade may indicate that members within this subfamily have the ability to synthesise either NANA or Leg. Further evidence for this is given by the minimal NANAS activity observed with *Cje*LegS (*64*). Although NANAS activity has not been directly investigated for *Lpn*LegS, complementation studies in *E. coli* have shown that *Lpn*NDP-sugar hydrolase/2-epimerase and *Lpn*LegS restored the biosynthesis of capsular polysialic acid in the corresponding *E. coli* K1 gene knock-out mutants. This strongly suggests that the SAS from *L. pneumophila* could synthesise either Leg (from 2,4-diacetamido-2,4, 6-trideoxymannose) or NANA (from *N*-acetylmannosamine) (*136*).

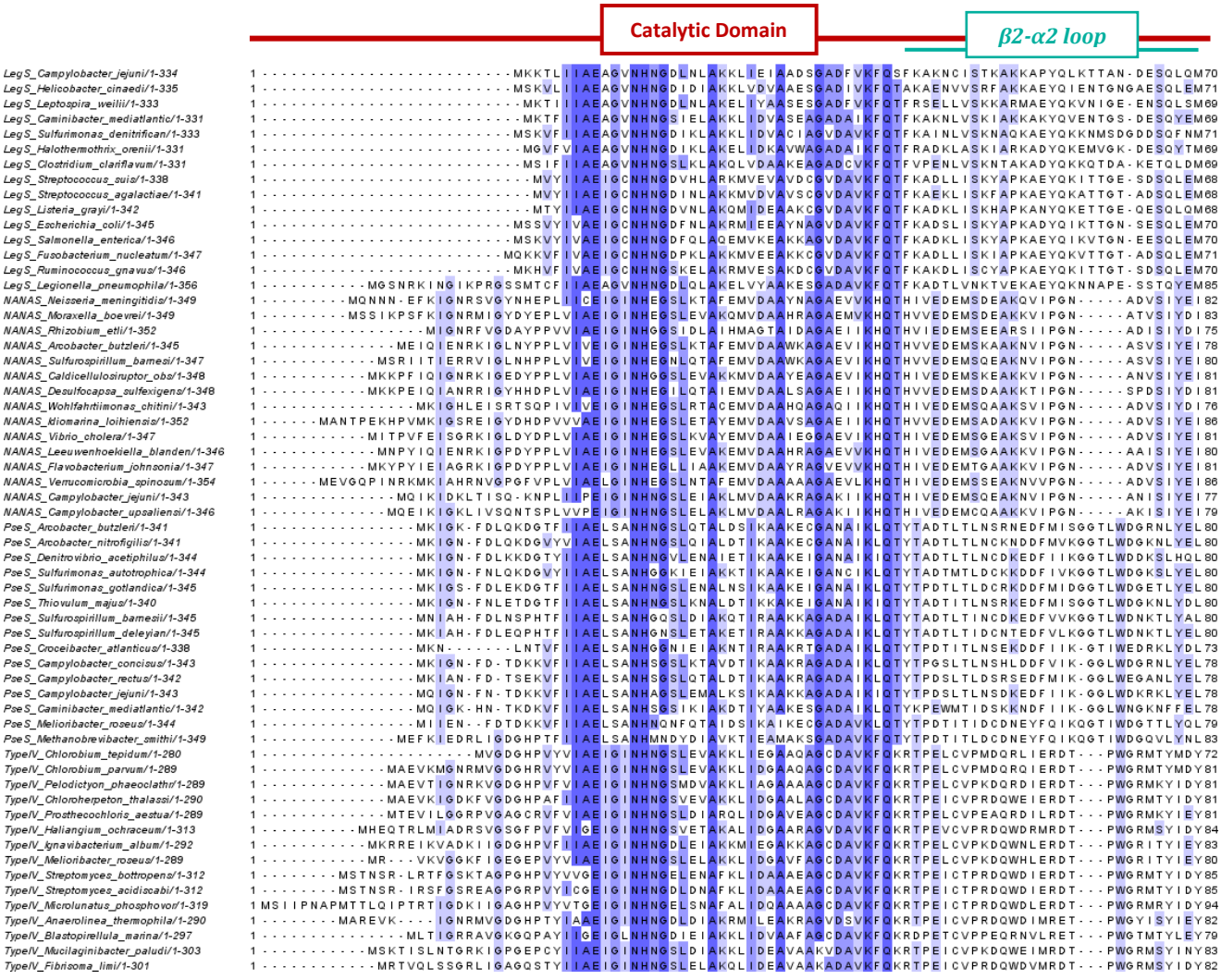
A third clade is denoted as the PseS subfamily. Within this grouping are both *Cje*PseS which has been kinetically characterised as a PseS (*89*), and *Hpy*PseS, which was used to elucidate the CMP-Pse pathway in *H. pylori (63)*. Additionally, this grouping contains the SASs from *P. profundum* strain SS9 and *Methanobrevibacter smithii*, which were both identified to produce Pse but not NANA or Leg through HPLC and MS (*134*). Unlike the LegS from *C. jejuni*, *Cje*PseS was reported to lack NANAS activity (*89*).

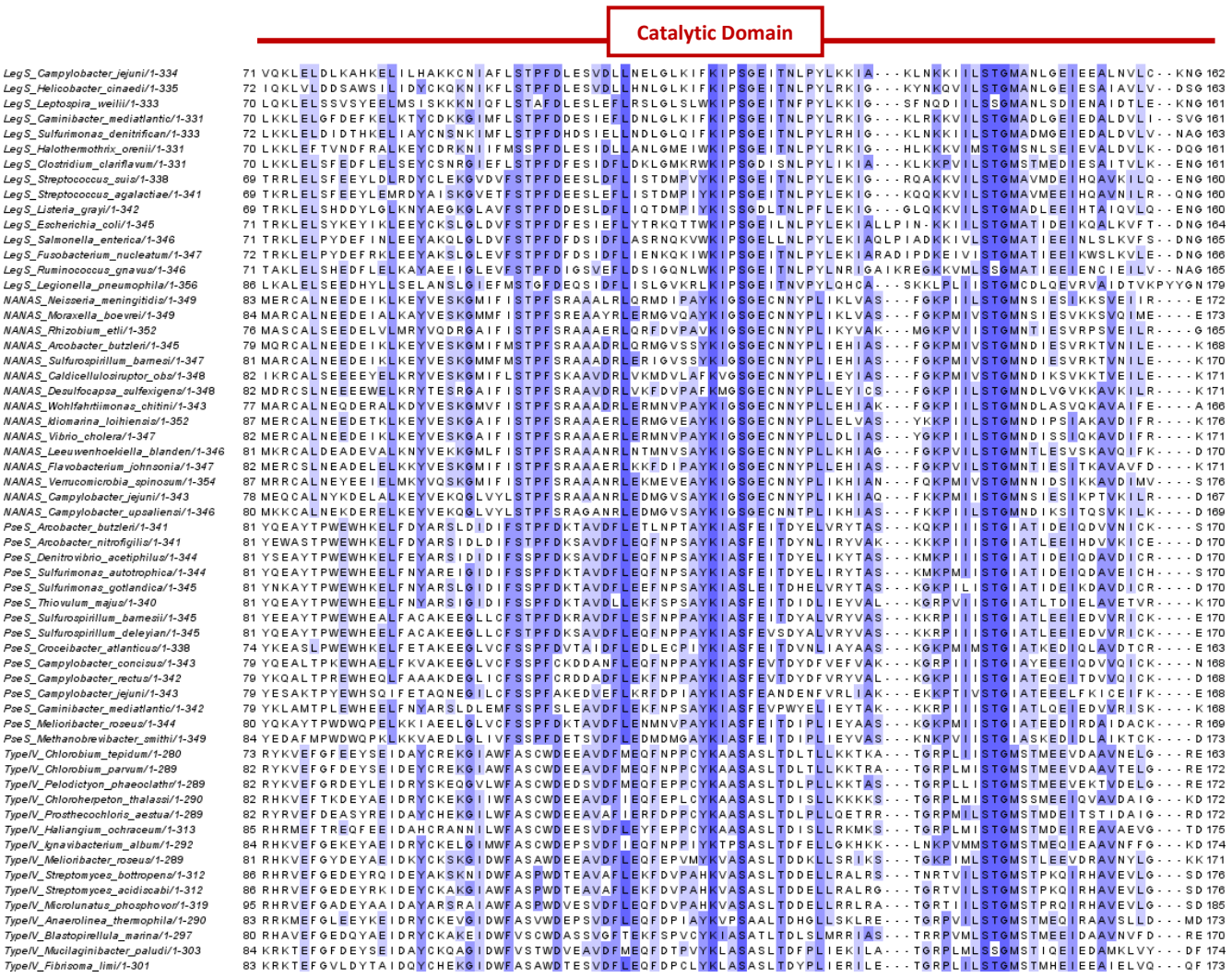
A fourth clade of SAS genes is identified in the bacterial SAS phylogram and is denoted as Type IV. Currently none of the enzymes from this particular subfamily have been tested for SAS activity in the literature. This grouping of unknown SAS enzymes has been previously noted in the literature (*134*). The phylogenetic analysis described by Lewis *et al* (2009), incorporates the mammalian SAS, NANA-9-PS. The SAS phylogram published in Lewis *et al* (2009) indicates that NANA-9-PS is more closely related to the unknown SAS enzymes than to any of the previously identified NANAS, LegS or PseS enzymes. In addition, certain members grouped in Type IV, such as the SAS from *Chlorobium tepidum*, have been previously hypothesised to have NANA-9-PS activity (*122*).

Lewis *et al* (2009) do not propose that the uncharacterised SAS orthologues are NANA-9-PSs. Instead they postulate that the origins of bacterial SAS enzymes did not arise through lateral gene transfer from the animal kingdom, but that NANA biosynthesis is a relic of an ancient pathway that was universally adopted by animals. This is evidenced by the distant sequence relationship observed between animal and 'animal-like' clades. Furthermore, the presence of archael sequences in each of the distinct SAS clades strongly supports the theory of paralogous gene duplication with divergence of enzymatic function occurring very early in cellular evolution (*134*). This is particularly interesting with respect to the AFPL domain and may suggest that the fold originated in the SAS family and was eventually commandeered and modified to resemble modern Type III anti-freeze proteins.

4.2.2 Sequence analysis of sialic acid synthases

To elucidate the key differences between the bacterial SASs that determines substrate selectivity, the previously mentioned SAS sequence alignment was thoroughly analysed. An abbreviated sequence alignment using 60 SAS sequences (15 from each clade) is shown in Figure 4.3. Active site residues determined from the substrate bound *Nme*NANAS crystal structure (1XUZ) (*112*) were identified in the complete sequence alignment, and the conservation of these residues is displayed in Table 4.1.




Catalytic Domain
LegS_Campylobacter_jejuni/1-334 71 VQKLELDLKAHKELILHAKKCNIAFLSTPFDLESVDLLNELGLKIFKIPSGEITNLPYLKKIA---KLNKKIILSTGMANLGEIEEALNVLC--KNG 162
LegS_Helicobacter_cinaedi/1-335 72 IQKLVLDDSAWSILIDYCKQKNIKFLSTPFDLESVDLLHNLGLKIFKIPSGEITNLPYLRKIG---KYNKQVILSTGMANLGEIESAIAVLV--DSG 163
LegS_Leptospira_weilii/1-333 70 LQKLELSSVSYEELMSISKKKNIQFLSTAFDLESLEFLRSLGLSLWKIPSGEITNFPYLKKIG---SFNQDIILSSGMANLSDIENAIDTLE--KNG 161
LegS_Caminibacter_mediatlantic/1-331 70 LKKLELGFDEFKELKTYCDKKGIMFLSTPFDDESIEFLDNLGLKIFKIPSGEITNLPYLRKIG---KLRKKVILSTGMADLGEIEDALDVLI--SVG 161
LegS_Sulfurimonas_denitrifican/1-333 72 LKKLELDIDTHKELIAYCNSNKIMFLSTPFDHDSIELLNDLGLQIFKIPSGEITNLPYLRHIG---KLNKKIILSTGMADMGEIEDALDVLV--NAG 163
LegS_Halothermothrix_orenii/1-331 70 LKKLEFTVNDFRALKEYCDRKNIIFMSSPFDLESIDLLANLGMEIWKIPSGEITNLPYLRKIG---HLKKKVIMSTGMSNLSEIEVALDVLK--DQG 161
LegS_Clostridium_clariflavum/1-331 70 LKKLELSFEDFLELSEYCSNRGIEFLSTPFDFESIDFLDKLGMKRWKIPSGDISNLPYLIKIA---KLKKPVILSTGMSTMEDIESAITVLK--ENG 161
LegS_Streptococcus_suis/1-338 69 TRRLELSFEEYLDLRDYCLEKGVDVFSTPFDEESLDFLISTDMPVYKIPSGEITNLPYLEKIG---RQAKKVILSTGMAVMDEIHQAVKILQ--ENG 160
LegS_Streptococcus_agalactiae/1-341 69 TKRLELSFEEYLEMRDYAISKGVETFSTPFDEESLEFLISTDMPIYKIPSGEITNLPYLEKIG---KQQKKVILSTGMAVMEEIHQAVNILR--QNG 160
LegS_Listeria_grayi/1-342 69 TRKLELSHDDYLGLKNYAEGKGLAVFSTPFDDESLDFLIQTDMPIYKISSGDLTNLPFLEKIG---GLQKKVILSTGMADLEEIHTAIQVLQ--ENG 160
LegS_Escherichia_coli/1-345 71 TRKLELSYKEYIKLEEYCKSLGLDVFSTPFDFESIEFLYTRKQTTWKIPSGELNLPYLEKIALLPIN-KKIILSTGMATIDEIKQALKVFT--DNG 164
LegS_Salmonella_enterica/1-346 71 TRKLELPYDEFINLEEYAKQLGLDVFSTPFDFDSIDFLASRNQKVWKIPSGELLNLPYLEKIAQLPIADKKIVLSTGMATIEEINLSLKVFS--DNG 165
LegS_Fusobacterium_nucleatum/1-347 72 TRKLELPYDEFRKLEEYAKSLGLEVFSTPFDFDSIDFLIENKQKIWKIPSGEITNLPYLEKIARADIPDKEIVISTGMATIEEIKWSLKVLE--DNG 166
LegS_Ruminococcus_gnavus/1-346 71 TAKLELSHEDFLELKAYAEEIGLEVFSTPFDIGSVEFLDSIGQNLWKIPSGEITNLPYLNRIGAIKREGKKVMLSSGMATIEEIENCIEILV--NAG 165
LegS_Legionella_pneumophila/1-356 86 LKALELSEEDHYLLSELANSLGIEFMSTGFDEQSIDFLISLGVKRLKIPSGEITNVPYLQHCA---SKKLPLIISTGMCDLQEVRVAIDTVKPYYGN 179
NANAS_Neisseria_meningitidis/1-349 83 MERCALNEEDEIKLKEYVESKGMIFISTPFSRAAALRLQRMDIPAYKIGSGECNNYPLIKLVAS---FGKPILSTGMNSIESIKKSVEIIR----E 172
NANAS_Moraxella_boewei/1-349 84 MARCALNEEDEIALKAYVESKGMMFISTPFSREAAYRLERMGVQAYKIGSGECNNYPLIKLVAS---FGKPMIVSTGMNSIESVKKSVQIME----E 173
NANAS_Rhizobium_etli/1-352 76 MASCALSEEDELVLMRYVQDRGAIFISTPFSRAAAERLQRFDVPAVKIGSGECNNYPLIKYVAK---MGKPVIISTGMNTIESVRPSVEILR----G 165
NANAS_Arcobacter_butzleri/1-345 79 MQRCALNEEDEIKLKEYVESKGMIFMSTPFSRAAADRLQRMGVSSYKIGSGECNNYPLIEHIAS---FGKPMIVSTGMNDIESVRKTVNILE----K 168
NANAS_Sulfurospirillum_barnesii/1-347 81 MARCALNEEDEIKLKEYVESKGMMFMSTPFSRAAADRLERIGVSSYKIGSGECNNYPLIEHIAS---FGKPMIVSTGMNDIESVRKTVNILE----K 170
NANAS_Caldicellulosiruptor_obs/1-348 82 IKRCALSEEEEYELKRYVESKGMIFLSTPFSKAAVDRLVKMDVLAFKVGSGECNNYPLIEYIAS---FGKPMIVSTGMNDIKSVKKTVEILE----K 171
NANAS_Desulfocapsa_sulfexigens/1-348 82 MDRCSLNEEEEWELKRYTESRGAIFISTPFSRAAVDRLVKFDVPAFKMGSGECNNYPLLEYICS---FGKPILSTGMNDLVGVKKAVKILR----K 171
NANAS_Wohlfahrtiimonas_chitini/1-343 77 MARCALNEQDERALKDYVESKGMVFISTPFSRAAADRLERMNVPAYKIGSGECNNYPLLEHIAK---FGKPILSTGMNDLASVQKAVAIFE----A 166
NANAS_Idiomarina_loihiensis/1-352 87 MERCALNEEDETKLKEYVESKGAIFISTPFSRAAAERLERMGVEAYKIGSGECNNYPLLELVAS---YKKPILSTGMNDIPSIAKAVDIFR----K 176
NANAS_Vibrio_cholera/1-347 81 MERCALNEEDEIKLKEYVESKGAIFISTPFSRAAAERLERMNVPAYKIGSGECNNYPLLDLIAS---YGKPILSTGMNDISSIQKAVDIFR----K 171
NANAS_Leeuwenhoekiella_blanden/1-346 81 MKRCALDEADEVALKNYVEKKGMLFISTPFSRAAANRLNTMNVSAYKIGSGECNNYPLLKHIAG---FGKPVILSTGMNTLESVSKAVQIFK----D 170
NANAS_Flavobacterium_johnsoniae/1-347 82 MERCSLNEADELELKKYVESKGMIFISTPFSRAAAERLKKFDIPAYKIGSGECNNYPLLEHIAS---FGKPVILSTGMNTIESITKAVAVFD----K 171
NANAS_Verrucomicrobia_spinosum/1-354 87 MRRCALNEYEEIELMKYVQSKGMIFISTPFSRAAANRLEKMEVEAYKIGSGECNNYPLIKHIAN---FQKPMIVSTGMNNIDSIKKAVDIMV----S 176
NANAS_Campylobacter_jejuni/1-343 78 MEQCALNYKDELALKEYVEKQGLVYLSTPFSRAAANRLEDMGVSAYKIGSGECNNYPLIKHIAQ---FKKPMIISTGMNSIESIKKPTVKILR----D 167
NANAS_Campylobacter_upsaliensi/1-346 80 MKKCALNEKDERELKEYVEKQGLVYLSTPFSRAGANRLEDMGVSAYKIGSGECNNTPLIKHIAS---FKKPIILSTGMNDIKSITQSVKILK----D 169
PseS_Arcobacter_butzleri/1-341 81 YQEAYTPWEWHKELFDYARSLDIDIFSTPFDKTAVDFLETLNPTAYKIASFEITDYELVRYTAS---KQKPIIISTGIATIDEIQDVVNICK----S 170
PseS_Arcobacter_nitrofigilis/1-341 81 YEWASTPWEWHKELFDYARSILDIFSTPFDKSAVDFLEQFNPSAYKIASFEITDYNLRYVAK---KKKPIIISTGIATLEEIHDVVKICE----D 170
PseS_Denitrovibrio_acetiphilus/1-344 81 YSEAYTPWEWHEELFEYARSIDIDIFSSPFDKSAVDFLEQFNPSAYKIASFEITDYELIKYTAS---KMKPIIISTGIATIDEIQDAVDICR----D 170
PseS_Sulfurimonas_autotrophica/1-344 81 YQEAYTPWEWHEELFNYARSIGIDIFSSPFDKTAVDFLEQFNPSAYKIASFEITDYELIRYTAS---KMKPMIISTGIATIDEIQDAVEICH----S 170
PseS_Sulfurimonas_gotlandica/1-345 81 YNKAYTPWEWHKELFNYARSLGIDIFSSPFDKTAVDFLEEFNPSAYKIASLEITDHELVRYTAS---KGKPILISTGIATIDEIKDAVDICR----D 170
PseS_Thiovulum_majus/1-340 81 YQEAYTPWEWHEELFNYARSIGIDIFSSPFDKTAVDLLEKFSPSAYKIASFEITDIDLIEYVAL---KGRPVIISTGIATLTDIELAVETVR----K 170
PseS_Sulfurospirillum_barnesii/1-345 81 YEEAYTPWEWHEALFACAKEEGLLCFSTPFDKRAVDFLESFNPPAYKIASFEITDYALVRYVAS---KKRPIIISTGIATLEEIEDVVRICK----E 170
PseS_Sulfurospirillum_deleyian/1-345 81 YQEAYTPWEWHEELFACAKEEGLLCFSTPFDKSAVDFLEQFNPPAYKIASFEVSDYALVRYVAS---KKRPIIISTGIATLEEIEDVVRICK----E 170
PseS_Croceibacter_atlanticus/1-338 74 YKEASLPWEWHKELFETAKEEGLVCFSSPFDVTAIDFLEDLECPIYKIASFEITDVNLIAYAAS---KGKPMIMSTGIATKEDIQLAVDTCR----E 163
PseS_Campylobacter_concisus/1-343 79 YQEALTPKEWHAELFKVAKEEGLVCFSSPFCKDDANFLEQFNPPAYKIASFEVTDYDFVEFVAK---KGRPIIISTGIAYEEEIQDVVQICK----N 168
PseS_Campylobacter_rectus/1-342 79 YKQALTPREWHEQLFAAAKDEGLICFSSPFCRDDADFLEKFNPPAYKIASFEVTDYDFVRYVAL---KGKPIIISTGIATEQEITDVVQICK----D 168
PseS_Campylobacter_jejuni/1-343 79 YESAKTPYEWHSQIFETAQNEGILCFSSPFAKEDVEFLKRFDPIAYKIASFEANDENFVRLIAK---EKKPTIVSTGIATEEELFKICEIFK----E 168
PseS_Caminibacter_mediatlantic/1-342 79 YKLAMTPLEWHEELFNYARSLDLEMFSSPFSLEAVDFLEKFNPSAYKIASFEVPWYELIEYTAK---KKRPIIISTGIATLQEIEDVVRISK----K 168
PseS_Melioribacter_roseus/1-344 80 YQKAYTPWDWQPELKKIAEELGLVCFSSPFDKTAVDFLENMNVPAYKIASFEITDIPLIEYAAS---KGKPMIISTGIATEEDIRDAIDACK----R 169
PseS_Methanobrevibacter_smithii/1-349 84 YEDAFMPWDWQPKLKKXVAEDLGLIVFSSPFDETSVDFLEDMDMGAYKIASFEITDIPLIEYVAS---KNKPVIISTGIASKEDIDLAIKTCK----D 173
TypeIV_Chlorobium_tepidum/1-280 73 RYKVEFGFEEYSEIDAYCREKGIAWFASCWDEEAVDFMEQFNPPCYKAASASLTDLTLLKKTKA---TGRPLIIISTGMSTMEEVDAAVNELG----RE 163
TypeIV_Chlorobium_parvum/1-289 82 RYKVEFGFDEYSEIDEYCREKGIAWFASCWDEEAVDFMEQFNPPCYKAASASLTDLTLLKKTRA---TGRPLMISTGMSTMEEVDAAVTELG----RE 172
TypeIV_Pelodictyon_phaeoclathr/1-289 82 RYKVEFGRDEYELIDRYSKEQGVLWFASCWDEDSVDFMEQFEPPCYKAASASLTDLPLLKKTAS---TGRPLLISTGMSTMEEVEKTVDELG----RE 172
TypeIV_Chloroherpeton_thalassi/1-290 82 RHKVEFTKDEYAEIDRYCKEKGIIWFASCWDEEAVDFIEQFEPLCYKAASASLTDISLLKKKKS---TGRPLMISTGMSSMEEIQVAVDAIG----KD 172
TypeIV_Prosthecochloris_aestua/1-289 82 RYRVEFDEASYREIDAYCHEKGILWFASCWDEEAVAFIERFDPPCYKAASASLTDLPLLQETRR---TGRPMILSTGMSTMDEITSTIDAIG----RD 172
TypeIV_Haliangium_ochraceum/1-313 85 RHRMEFTREQFEEIDAHCRANNILWFASCWDEESVDFLEYFEPPCYKAASASLTDISLLRKMKS---TGRPLMISTGMSTMDEIREAVAEVG----TD 175
TypeIV_Ignavibacterium_album/1-292 84 RHKVEFGEKEYAEIDRYCKELGIMWFASCWDEPSVDFIEQFNPPIYKTPSASLTDFELLGKHKK---LNKPVMMSTGMSTMEQIEAAVNFFG----KD 174
TypeIV_Melioribacter_roseus/1-289 81 RHKVEFGYDEYAEIDKYCKSKGIDWFASAWDEEAVDFLEQFEPVMYKVASASLTDDKLLSRIKS---TGKPIMLSTGMSTLEEVDRAVNYLG----KK 171
TypeIV_Streptomyces_bottropens/1-312 86 RHRVEFGEDEYRQIDEYAKSKNIDWFASPWDTEAVAFLEKFDVPAHKVASASLTDDELLRALRS---TNRTVILSTGMSTPKQIRHAVEVLG----SD 176
TypeIV_Streptomyces_acidiscabi/1-312 86 RHRVEFGEDEYRKIDEYCKAKGIAWFASPWDTEAVAFLEKFDVPAHKVASASLTDDELLRALRG---TGRTVILSTGMSTPKQIRHAVEVLG----SD 176
TypeIV_Microlunatus_phosphovor/1-319 95 RHRVEFGADEYAAIDAYARSRAIAWFASPWDVESVDFLEQFDVPAHKVASASLTDDELLRRLRA---TGRTIILSTGMSTPRQIRHAVEVLG----SD 185
TypeIV_Anaerolinea_thermophila/1-290 83 RRKMEFGLEEYKEIDRYCKEVGIDWFASVWDEPSVDFLEQFDPIAYKVPSAALTDHGLLSKLRE---TGRPVILSTGMSTMEQIRAAVSLLD----MD 173
TypeIV_Blastopirellula_marina/1-297 80 RHAVEFGEDQYAEIDRYCKAKEIDWFVSCWDASSVGFTEKFSPVCYKIASATLTDLSLMRRIAS---TRRPVMLSTGMSTMEEIDAAVNVFD----RE 170
TypeIV_Mucilaginibacter_paludi/1-303 84 KRKTEFGFDEYSEIDAYCKQAGIDWFVSTWDVEAVDFMEQFDTPVYKLASASLTDFPLIEKILA---TGRPLMLSSGMSTIQEIEDAMKLVY---DF 174
TypeIV_Fibrisoma_limi/1-301 83 KRKTEFGVLDYTAIDQYCHEKGIDWFASAWDTESVDFLEQFDPCLYKLASASLTDYPLIERILE---TGRPLILSTGMSTMHEIEEAIELVQ---QF 173
Catalytic Domain
LegS_Campylobacter_jejuni/1-334 163 AKRQNITLLHCTTEYPAPFNEVNLKAMQSLKDAFK-LDVGYSDHTRGIHISLAAVALGACVIEKHFTLDKNMSGPDHKASLEPQELKMLCTQIRQIQ 258
LegS_Helicobacter_cinaedi/1-335 164 TKRENITLLQCNTEYPTPFADVNLKAMKSLKKAFR-LPVGYSDHTPGIAIPLAAVGMGAKVIEKHFTLDKNMEGPDHKASLEPCELKAMVQGIREIE 259
LegS_Leptospira_weilii/1-333 162 TKREKMTVLHCNTEYPTPFEDVNLNAMNTIRSAFH-TKVGYSDHTSGIEVAIAAVALGATVIEKHFTLDRNMSGPDHKASLEPSELKRMVSSIRNIE 257
LegS_Caminibacter_mediatlantic/1-331 162 IKKKNITVLHANTEYPTPIEDVNLKAMKTIECAFG-VDVGYSDHTLGIEVPIAAVAMGAKVIEKHFTLSRDLEGPDHKASLEPDELKEMIKSIRNIE 257
LegS_Sulfurimonas_denitrifican/1-333 164 TKKENITVLHANTMYPTPMEDVNLRAMITIGEAFD-VAFGYSDHTLGIEVPIAAVAMGATCIEKHFTLDCTMDGPDHKASLEPDELKAMVKAIRNIE 259
LegS_Halothermothrix_orenii/1-331 162 VN--DITVLHCNTEYPTPIEDVNLKAMQTIQNAFK-VKVGYSDHTRGIEVPIAAVALGAEVIEKHFTLDRDMEGPDHKASLEPEELKIMIEFIRNIE 255
LegS_Clostridium_clariflavum/1-331 162 AD--DITVLHCNTEYPTSYSDVNLKAMLSIKEKFN-VSIGYSDHTKGIEVSIAAALGASVIEKHFTLDSRMEGPDHKASLEPNELKAMVDAIRNIE 255
LegS_Streptococcus_suis/1-338 161 TT--DISILHCTTEYPTPYPALNLNVLHTLKKEFPNLTIGYSDHSVGSEVPIAAAAMGAELIEKHFTLDNEMEGPDHKASATPDILAALVKGVRIVE 255
LegS_Streptococcus_agalactiae/1-341 161 TT--DISILHCTTEYPTPYPSLNLNVIHTLKDEFKDLTIGYSDHSIGSEVPIAAAAMGAEVIEKHFTLDTNMEGPDHKASATPDILAALVKGVRIVE 255
LegS_Listeria_grayi/1-342 161 TS--DIALLHCTTEYPTAFPDVNLQVLRTFQTLFPDLEIGYSDHSLGYEVPIAAVAMGAQVIEKHFTLDNLDPGPDHRASATPDILKRLVEGIRHIE 255
LegS_Escherichia_coli/1-345 165 INSEDITILHCNTEYPTPYEDVNLNALAGLKKTFQGYNIGFSDHSLGPGFFAGIAAVPYGITFIEKHFTLNKNFEGPDHKASVTPDELRLLCEGIRAIE 261
LegS_Salmonella_enterica/1-346 166 ISPENITILHCNTEYPTPFEDVNLNTICGFKKIFEQYNIGFSDHSPGYFAGIASVPYGITFIEKHFTLDKNLEGPDHKASVTPDELKLLCEGIRAVE 262
LegS_Fusobacterium_nucleatum/1-347 167 VSKNKITILHCNTEYPTPYEDVNLEAIPYLMEKFPEYKIGFSDHSQGYFDGVASVIYGINFIEKHFTLDKNFEGPDHKASVTPDELKLLCEGIRVVE 263
LegS_Ruminococcus_gnavus/1-346 166 TKECDIILLHCNTEYPTDKDVNILAINDLKKHFPNVKIGFSDHSVGYTAAIGAAVLGIDLIEKHFTLDKNYEGPDHKASATPNELKELVQNVRRVE 262
LegS_Legionella_pneumophila/1-356 180 SLSDYLVLLHCTSNYPASYQDVNLKAMQTLADEFQ-LPVGYSDHTLGILVPTLAVGMGACVIEKHFTMDKSLPGPDHLASMDPEEMKNLVQSIRDAE 275
NANAS_Neisseria_meningitidis/1-349 173 A-GVPYALLHCTNIYPTPYEDVRLGGMNDLSEAFPDAIIGLSDHTLDNYACLGAVALGGSILERHFTDRMDRPGPDIVCSMNPDTFKELKQGAHALK 268
NANAS_Moraxella_boevrei/1-349 174 A-GVPYALLHCTNVYPTPPELVRLGAMNELEEAFPNAVIGLSDHTCIDNYVCMGAVALGASILERHFTDSFDRPGPDINCSMDSKGMAELIKASRQIA 269
NANAS_Rhizobium_etli/1-352 166 A-GVPYALLHCTNVYPTPPELVRLGAMASLREAFPDAVIGLSDHTTSNYTCLGAVALGASILERHFTDSMSREGPDIVCSMDPPALSDLIEGSRTLF 261
NANAS_Arcobacter_butzleri/1-345 169 Y-NIPYALLHTTNLYPTPVHLVRLGAMQELQKEFPNAIIGLSDHTTSNRACFAATALGACILERHFTDKMERPGPDIINSMDPIALKELIYGSAEIA 264
NANAS_Sulfurospirillum_barnesi/1-347 171 Y-KVPYALLHTTNLYPTPVHLVRLGAMQELQKAFPNAVVGLSDHTTSNRACFAATALGASILERHFTDKMERPGPDIINSMDPIALKELIIGSNEIA 266
NANAS_Caldicellulosiruptor_obs/1-348 172 Y-NIQYALLHCTNLYPTPPHLVRLGGLLELKETFPNAVIGLSDHTVNNYACLAAVALGASILERHFTDSKDRPGPDIVCSMDPKELKELIEASAQIA 267
NANAS_Desulfocapsa_sulfexigens/1-348 172 H-TIPFALLHTTNLYPTPPELVRLGALLDLEEAFPDAVIGLSDHTTNNLACLGAVALGASILERHFTDSKDRPGPDIVCSMTPSELNELIASSKILK 267
NANAS_Wohlfahrtiimonas_chitini/1-343 167 H-GVPYALLHTTNLYPTPAHLVRLGAMEVLADAFVNAPFGLSDHTTSNLACFAAVALGATLLERHFTDHKSRPGPDIINSMDAQELSELIHGSRLIH 262
NANAS_Idiomarina_loihiensis/1-352 177 H-GVQYALLHCTNVYPTPDNLIRLGAMNQLQDAFDDAVVGLSDHSIDNLACLGAVAVGASILERHFTDHKQRTGPDINCSMDTEECRSLIQESKRMK 272
NANAS_Vibrio_cholera/1-347 172 H-KTPFCLLHTTNLYPTPDHLIRIGAMEQLQDAFPDAVVGLSDHSIDNLACLGAVAAGASVLERHFTDRKDRPGPDIVCSMDAKECAELIAQSKRMA 267
NANAS_Leeuwenhoekiella_blanden/1-346 171 A-GTPFALLHTTNLYPTPHLVRLGAMTQLHHAFPDVPFGLSDHTLNNNACLGAVALGASILERHFTDHMQRTGPDIVCSMDEEACAELIQSSKEIW 266
NANAS_Flavobacterium_johnsonia/1-347 172 H-NIPVALLHTTNLYPTPIHLVRFGAMLELHQAFPDKVFGLSDHTLNNNACLGAVALGASILERHFTDHMQRTGPDIVCSMDEEACAELIISSAEIA 267
NANAS_Verrucomicrobia_spinosum/1-354 177 N-EVNFALLHTTNIYPTPNHLVRLHAVREMMDAFPGVIVGLSDHTLSNHACFGAVALGCSIVERHFTDHMGREGPDIVCSMDQERLKELIDGVEILF 272
NANAS_Campylobacter_jejuni/1-343 168 Y-EIPFVLLHTTNLYPTPSHLVRLQAMLELYKEFN-CLYGLSDHTTNNLACIGAIALGASVLERHFTDTMDRKGPDIVCSMDESTLKDLINQTQEMV 262
NANAS_Campylobacter_upsaliensi/1-346 170 F-EVPFVLLHTTNLYPTPHLVRLHAMLTLQKEFN-ALVGLSDHTTNNLACLGAVALGACMIERHFSDTMDRQGPDIICSMDENALKELITQSEQMA 264
PseS_Arcobacter_butzleri/1-341 171 VGNENIILLKCTSEYPAPLNEANLKTIPDMKERFN-VEVGFSDHTLGIAVPVVAVTLGAKVIEKHFIIDKSIGGADCGFSMDKKEFKQMVEAIRDAE 266
PseS_Arcobacter_nitrofigilis/1-341 171 EGNEDIILLKCTSSYPAPLEEANLKTIPDMKERFN-VEVGFSDHTLGSTAPIIAVAYGAKIIEKHFILDKSVGGADAEFSMEKDEFSFMVKAIRESE 266
PseS_Denitrovibrio_acetiphilus/1-344 171 VGNNEIVLLKCTSAYPAPLEDANLITINNLSDTFG-VISGFSDHTLGITAPIVATTLGAKVIEKHFILDKAIGGADSDFSLDKDEFKDMIKAVRDAE 266
PseS_Sulfurimonas_autotrophica/1-344 171 VDNNEIVLLKCTSAYPAALEDANLLTIKNLAETFD-VVSGFSDHTMGATAPIVATLGAKVIEKHFILDKSIGGADADFSMDKEEFKAMIQAVRDTE 266
PseS_Sulfurimonas_gotlandica/1-345 171 AGNNDIILLKCTSSYPAPLEDANLLKTIPNLAETFD-VISGFSDHTLGSIAPIVAVALGAKVIEKHFILDKEIGGADAKFSMDKNDFARMINDIRNAE 266
PseS_Thiovulum_majus/1-340 171 TGNNDIILLKCTSSYPAPLEEINLNTIPNLKETFG-VTVGLSDHTLGISVPVGAVALGAKIVEKHFIADRKIGGPDSAFSLDPKQFKNMINDSVREIE 266
PseS_Sulfurospirillum_barnesii/1-345 171 EGNEEIVLLQCTSSYPAPLEDANLLMIPELQKRFG-VIAGFSDHTLGLTAPVMAVALGAKVIEKHFILDKSIGGADASFSLDVEAFSAMVHAVREAE 266
PseS_Sulfurospirillum_deleyian/1-345 171 EGNEEIVLLQCTSAYPAPLENANLLMIPDLQERFG-VVAGFSDHTLGITAPVMAVALGAKVIEKHFILDKSIGGADASFSLDVEAFSQMSKAVREAE 266
PseS_Croceibacter_atlanticus/1-338 164 AGNNDITLLKCTSSYPAPIDEANLVMMQRFAKDFN-VKVGLSDHTLGTTLPVVATALGAKVIEKHFILDKSIGGPDASFSLDEEEFTQMVSEVRNAE 259
PseS_Campylobacter_concisus/1-343 169 VGNNDIALLKCTSSYPAPLNGMNLQTIANMREKFG-VEVGFSDHTLGVTAPVVAVSLGARIVEKHFILDKSVKSVDSAFSLDESEFALMIKCVREAE 264
PseS_Campylobacter_rectus/1-342 169 AGNERIALLKCTSSYPAPLDGMNLRTIAGMKRKFG-VEVGFSDHTLGIVAPVVAVSLGARIVEKHFILDKSVRSVDEAFSLDEREFSEMVKAVRDAE 264
PseS_Campylobacter_jejuni/1-343 169 EKNPDLVFLKCTSTYPTAIEDMNLKGIVSLKEKFN-VEVGLSDHSFGFLAPVMAVALGARVIEKHFMLDKSIESEDSKFSLDFDEFKAMVDAVRQAE 264
PseS_Caminibacter_mediatlantic/1-342 169 E-NVEIALLKCTSAYPTPLEDANLKTIPAIKEIFN-VEVGFSDHTLGVTAPIVAVSLGARIIEKHFILDKNLNSPDKEFSLTPNEFKEMVKTIREAE 263
PseS_Melioribacter_roseus/1-344 170 MNNDQIALLKCTSSYPAPIEDANLNTIPYMKEKFN-TAVGLSDHTMGTVAPVVAVALGARIIEKHFILDRNLGGPDSAFSLEPAEFKNMVEGIRNAE 265
PseS_Methanobrevibacter_smithii/1-349 174 AGNDKIAVLKCTSSYPAPLEEINLKTIPDLKENFK-TVVGLSDHTLGSDVAVASVAMGAKIIEKHFILDRTMEGPDSDFSMEPDEFKQMVDSIRNVE 269
TypeIV_Chlorobium_tepidum/1-280 164 N----LLIAHTNSTYPCPVEELNLRMIHTLQQRYPDNPIGYSGHEVGLATTWAAVALGATFIERHVTLDRAMWGSDQAASVEISGMSRLVSNIDIE 256
TypeIV_Chlorobium_parvum/1-289 173 N----LLIAHTNSTYPCPVEELNLRMIHTLQQRYPDNPIGYSGHEVGLATTWAAVALGATFIERHVTLDRAMWGSDQAASVEISGMARLVANIRDIE 265
TypeIV_Pelodictyon_phaeoclathr/1-289 173 N----ILIAHTNSTYPSPIEELNLRMITTLKARYPEVPIGYSGHEVGLATTWAAVALGATFIERHVTLDRAMWGSDQAASVELPGMARLVSNIRDIE 265
TypeIV_Chloroherpeton_thalassi/1-290 173 N----LLIAHSTSTYPCPVKELNLRMIQTLKGLYPECPIGYSGHEVGLSTTWASVALGATFIERHITLDRAMWGTDQAASVEVGGMMRLVSNTRDIE 265
TypeIV_Prosthecochloris_aestua/1-289 173 N----LLLAHSNSTYPCPVEELNLRMIHSLQEMYPDIPVGYSGHEVGLATTWAAVALGATFIERHVTLDRAMWGSDQAASVELAGLKRLVTNIRDIE 265
TypeIV_Haliangium_ochraceum/1-313 176 N----LLLAHSTSAYPCSTKELNLRVIPNLAHMFPECPIGYSGHEEGITPTLSAVTLGATFVERHITLDCSMWGSDQAASVEPGMLSELVRNIREVE 268
TypeIV_Ignavibacterium_album/1-292 175 N----LLLAHATSAYPCKNDELNLRMIETLRKKYPDIPIGYSGHEVGLAPTWAAVALGACFVERHITLDRAMWGTDQAASVEMGGLQRLVSNIRDIE 267
TypeIV_Melioribacter_roseus/1-289 172 N----LLIAQSTSSYPCSLEELNLRVIDTYKKRYPEIPIGYSGHETGLAPSLAAVALGAVFVERHITLDRAMWGTDQAASVEIGGMYRLVKDIRDIE 264
TypeIV_Streptomyces_bottropens/1-312 177 N----LLLCHATSTYPAKAEELNLRVINTLQAEYPNVPIGYSGHETGLQTTLAAVALGAAFVERHITLDRAMWGSDQAASVEPGGLTRLVRDIRTIE 269
TypeIV_Streptomyces_acidiscabi/1-312 177 N----LLLCHATSTYPAKAEELNLRVINTLQAEYPNVPIGYSGHETGLQTTLAAVALGAAFVERHITLDRAMWGSDQAASVEPGGLTRLVRDIRTIE 269
TypeIV_Microlunatus_phosphovor/1-319 186 N----IVLCHATSTYPAKPLELNLRMIHTLQADFPNVPIGYSGHEVGLQTTLAAVALGATFVERHITLDRAMWGSDQAASVEPGLQRLVRDIRTIT 278
TypeIV_Anaerolinea_thermophila/1-290 174 K----LVITHATSTYPCDPEELNLRVIEKLRAEFP-CPIGYSGHEVGLIPSVVAVALGACLVERHITLDRAMWGSDQAASVEPQGFERLVKYIRVTE 265
TypeIV_Blastopirellula_marina/1-297 171 Q----LLIAHATSAYPCKAEEINLRMIHVLRDTYADVPIGYSGHEVGLQITLAAVALGAAFVERHITLDRTMWGSDHAASVEPQGLIRLCRDIRVIE 263
TypeIV_Mucilaginibacter_paludi/1-303 175 NPEYPVMLAHSTSAYPCKPEELNLKMIPTLSAKYPEAPIGYSGHETGLATTVAAAVMGATFVERHFTLDRAMWGSDHAASVEPQGFHRLVRDIRDIE 271
TypeIV_Fibrisoma_limi/1-301 174 DADYPLLVAHSTSAYPCKPEELNLRMIPTLQSRYPNLPIGYSGHETGLATTVAAVTLGATFIERHFTLDRAMWGSDHAASVEPQGLQRLVRDIRDVE 270

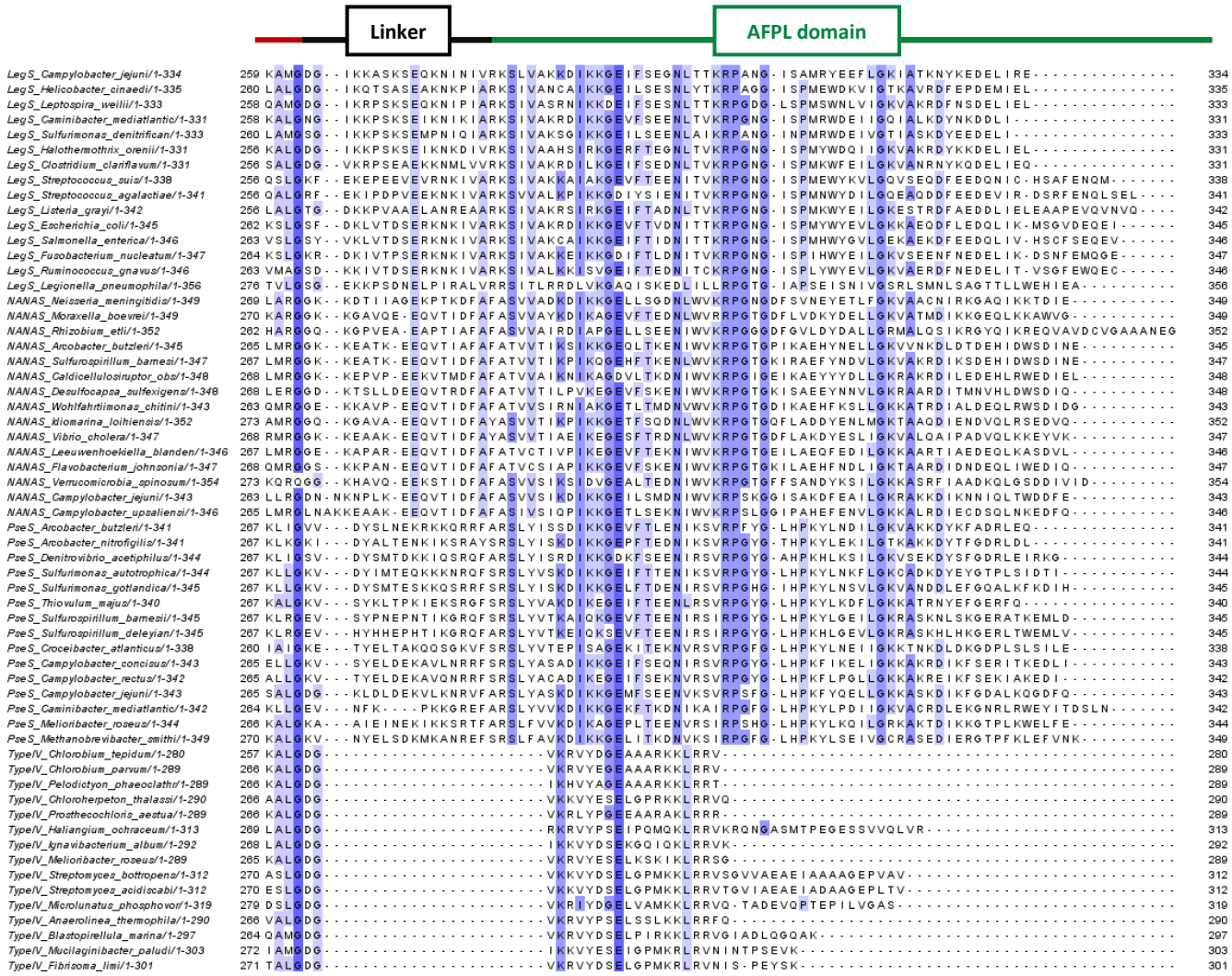


Figure 4.3 Alignment of 60 SAS sequences displayed using Jalview (*159*). 15 sequences were selected from each of the four SAS clades. Residues highlighted in blue indicate conservation across the SAS family, with increasing intensity of blue indicating higher identity and absence of blue indicating lack of identity. The domain annotation is based on the *Nme*NANAS crystal structure (PDB code (1XUZ) (*112*)).

Residue	Domain	Proximal ligand	Contact to ligand	Conserved in all bacterial SASs	Conserved in all NANASs
His215	Catalytic	Mn^{2+}	Dipolar	Yes	Yes
His236	Catalytic	Mn^{2+}	Dipolar	Yes	Yes
Lys053	Catalytic	PEP	H-bond	Yes	Yes
Thr110	Catalytic	PEP	None	Thr or Ser only	Yes
Lys129	Catalytic	PEP	H-bond	Yes	Yes
Ser132	Catalytic	PEP	H-bond	Yes	Yes
Ser154	Catalytic	PEP	H-bond	Yes	Yes
Ser213	Catalytic	PEP	H-bond	Yes	Yes
Asn184	Catalytic	PEP	H-bond	No	Yes
Gln055	Catalytic	rManNAc	H-bond	Yes	Yes
Asn074	Catalytic	rManNAc	H-bond	No	Yes
Tyr186	Catalytic	rManNAc	H-bond	Yes	Yes
Asp247	Catalytic	rManNAc	H-bond	Yes	Yes
Phe288	Linker	rManNAc	None	No	Yes
Arg314	AFPL	rManNAc	H-bond	Except Type IV	Yes

Table 4.1 Conserved SAS residues. Residues are numbered according to the sequence from *Nme*NANAS. Ligand proximity and Hydrogen bonding contacts are determined with respect to the crystal structure of *Nme*NANAS bound with Mn^{2+}, PEP and rManNAc (PDB code 1XUZ).

The members of the SAS family have relatively high sequence identity (98.7 % to 21.1 %) and are likely to share a similar overall fold to what is observed in the *Nme*NANAS crystal structure (Figure 4.4A). Observed in Figure 4.3, a notable difference between the various clades of the SAS family is that unlike NANAS, LegS and PseS enzymes, Type IV SASs do not contain peptide sequence pertaining to a C-terminal AFPL domain.

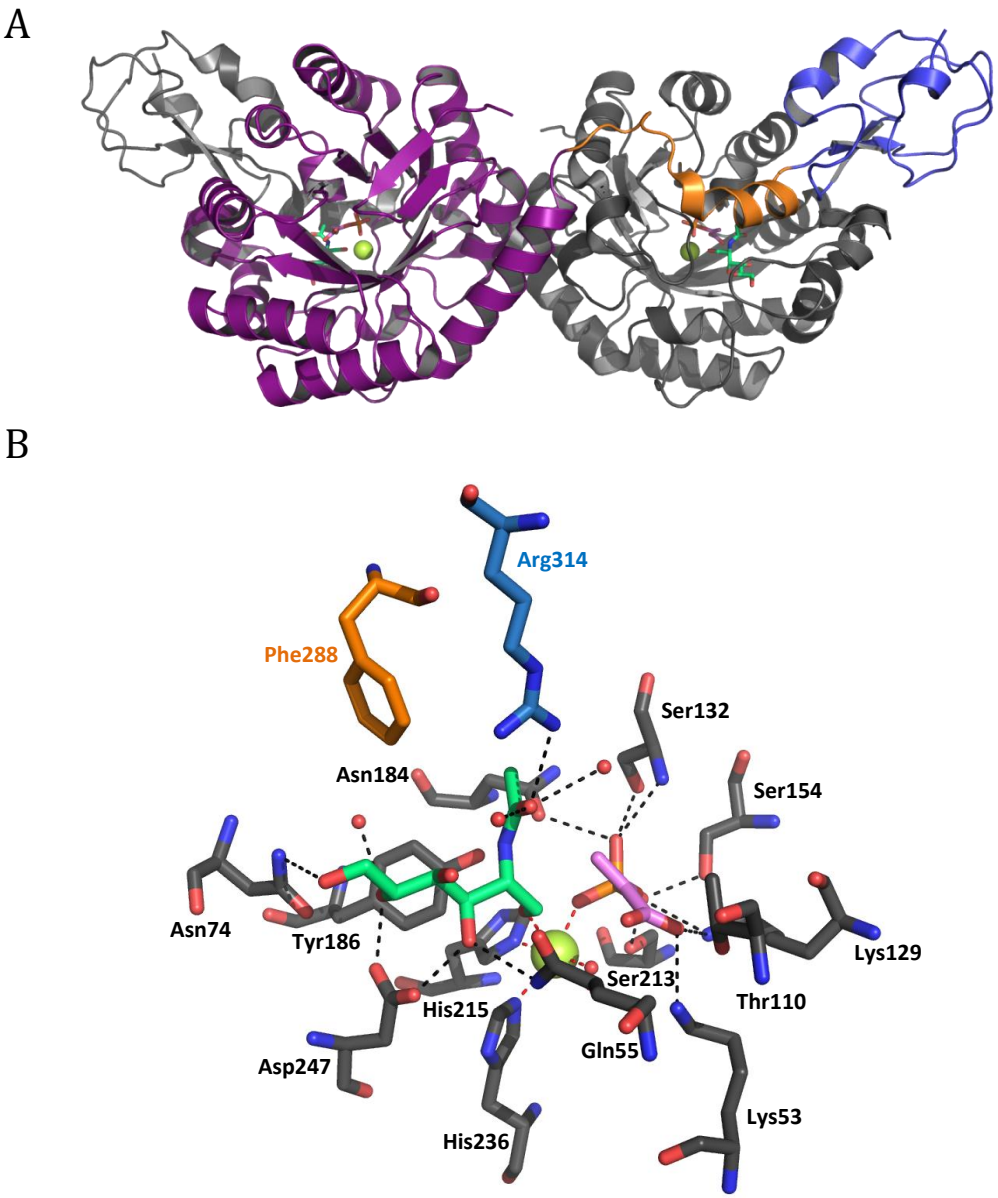


Figure 4.4 (A) *Nme*NANAS dimer with one monomer coloured grey and the second coloured by domain. The catalytic domain is shown in purple, linker domain in orange and AFPL domain in blue. (B) Close up of active site. Residues contributed by the catalytic domain are in grey, linker in orange and AFPL domain in blue. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are shown in red. H-bonds are depicted with black dashes and dipolar bonds with red dashes (PDB code (1XUZ) (*112*)).

Each of the SAS enzymes currently described in the literature has required a divalent metal ion for catalysis (*5, 89*). As shown in Figure 4.3 and Table 4.1, the metal binding residues ^{Nme}His215 and ^{Nme}His236 are both fully conserved in all members of the SAS family. The metal ion in the *Nme*NANAS crystal structure is shown to have octahedral geometry with ^{Nme}His215, ^{Nme}His236, a water molecule and O2P from PEP occupying the equatorial plane whilst a second water molecule and the carbonyl group of ManNAc occupy the axial positions (Figure 4.4B). ^{Nme}His215 and ^{Nme}His236 are the only metal binding ligands that are contributed by the enzyme itself, and the full conservation of these residues within the SAS family strongly indicates that all SASs are metal-dependent enzymes.

The phosphate group of PEP forms hydrogen bonds with ^{Nme}Ser132, ^{Nme}Ser154 and ^{Nme}Ser213 in the crystal structure of *Nme*NANAS. Each of these three serine residues is fully conserved within the SAS family. ^{Nme}Asn184 is also involved in PEP phosphate binding, however, it is only fully conserved in enzymes from the NANAS clade (Figure 4.3). In both PseS and Type IV SASs, this position correlates with a fully conserved serine, and in LegS a highly conserved threonine. Residues ^{Nme}Lys53 and ^{Nme}Lys129 form favourable interactions with the PEP carboxylate group, and are both conserved in all SAS enzymes. The side-chain of ^{Nme}Thr110 is proximal to the PEP carboxylate, but is not close enough to form a hydrogen bond in the *Nme*NANAS crystal structure. The positioning of ^{Nme}Thr110 is consistent amongst available *Nme*NANAS structures (*113*). ^{Nme}Thr110 is highly conserved as either a threonine or serine (Figure 4.3).

In the *Nme*NANAS crystal structure several strong hydrogen bonding interactions with rManNAc are noted (*112*). Unlike the metal and PEP binding sites, the ManNAc binding site is composed of residues from both the catalytic domain and the AFPL domain (Figure 4.4B). The catalytic domain residues ^{Nme}Asp247, ^{Nme}Gln55 and ^{Nme}Tyr186 are each fully conserved in the SAS family and form hydrogen bonds with rManNAc. ^{Nme}Asn74 is also contributed by the catalytic domain but is only conserved in the NANAS clade. In the *Nme*NANAS crystal structure, ^{Nme}Asn74 interacts with the C-6 hydroxyl of rManNAc. As shown in Figure 4.1, the substrates of both PseS and LegS do not contain a C-6 hydroxyl and thus conservation of ^{Nme}Asn74 would be unnecessary from an evolutionary perspective.

^{Nme}Phe288 and ^{Nme}Arg314 are the only two catalytic residues contributed by the linker and AFPL domain (Table 4.1) and are thus not present in Type IV SASs. ^{Nme}Arg314 interacts with the *N*-acetyl group at C-2 of ManNAc and is involved in steering the sugar substrate into a suitable reactive position (*114*). ^{Nme}Arg314 is conserved in all NANAS, LegS and PseS enzymes. As shown in Figure 4.1, the *N*-acetyl group at position C-2 is present in the sugar substrates of NANAS, LegS and PseS enzymes. It is thus highly likely that the positioning role of ^{Nme}Arg314 is maintained in these three SAS clades. Unlike ^{Nme}Arg314, ^{Nme}Phe288 does not directly interact with ManNAc but is in close proximity and forms part of the ManNAc binding pocket (Figure 4.4B). ^{Nme}Phe288 is only conserved in members of the NANAS clade.

There are two prominent regions where considerable sequence variation is observed within the entire SAS family but high conservation is observed within members of the same clade. These regions are likely indicative of residues, combination of residues or motifs that enable the SAS enzymes to differentiate between the required sugar substrates. The first of these regions is the β2-α2 loop of the catalytic domain shown in Figure 4.5. In the crystal structure of *Nme*NANAS, the β2-α2 loop closes over the ManNAc site and contributes the residue ^{Nme}Asn74, which forms a hydrogen bond with the C-6 hydroxyl of ManNAc. ^{Nme}Asn74 is fully conserved in all NANASs but is not present in any other SAS.

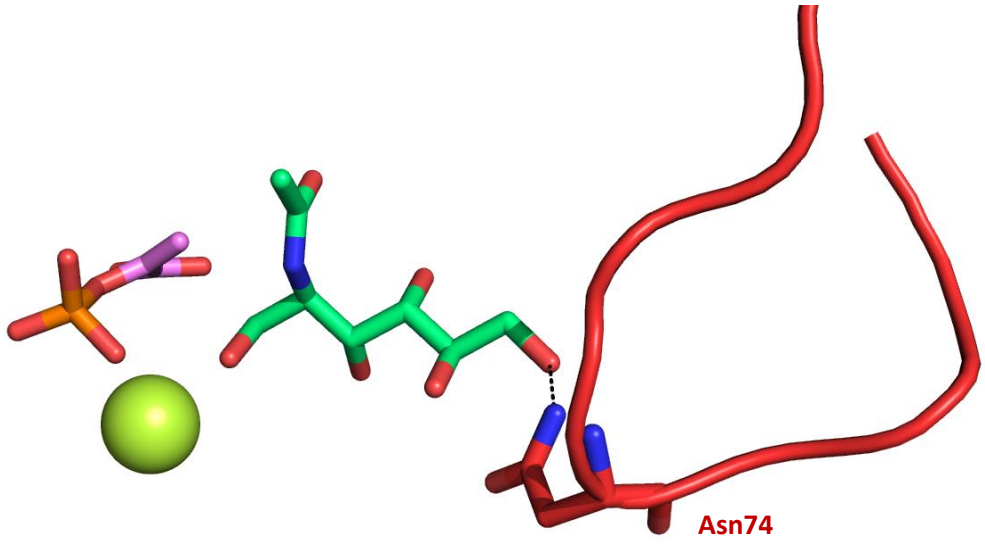


Figure 4.5 View of catalytic site and β2-α2 loop. β2-α2 loop is coloured in red. ManNAc is coloured green, PEP pink and Mn^{2+} yellow. H-bonds are depicted with black dashes (PDB code (1XUZ) (*112*)).

As can be seen in Figure 4.6, the residues comprising the β2-α2 loop differ significantly within the entire SAS family but show remarkably high conservation within each clade. Furthermore, as depicted in Figure 4.6, there are several residues that are fully conserved within each of the four SAS clades. In the NANAS clade the residues surrounding ^{Nme}Asn74 are fully conserved, resulting in a Pro-Gly-Asn motif. Similarly, in the LegS clade there is a cluster of conserved residues in the β2-α2 loop resulting in an Ala-X-Tyr-Gln motif, where the residue indicated by X is variable. In the PseS clade, there are five fully conserved residues within the β2-α2 loop region, a tyrosine, threonine, phenylalanine, tryptophan and a leucine. Unlike the NANAS and LegS clades, the conserved residues in the PseS β2-α2 loop are sparsely distributed. Lastly, in the Type IV SAS clade, there is a conserved Arg-X-Thr-Pro-Trp-Gly motif.

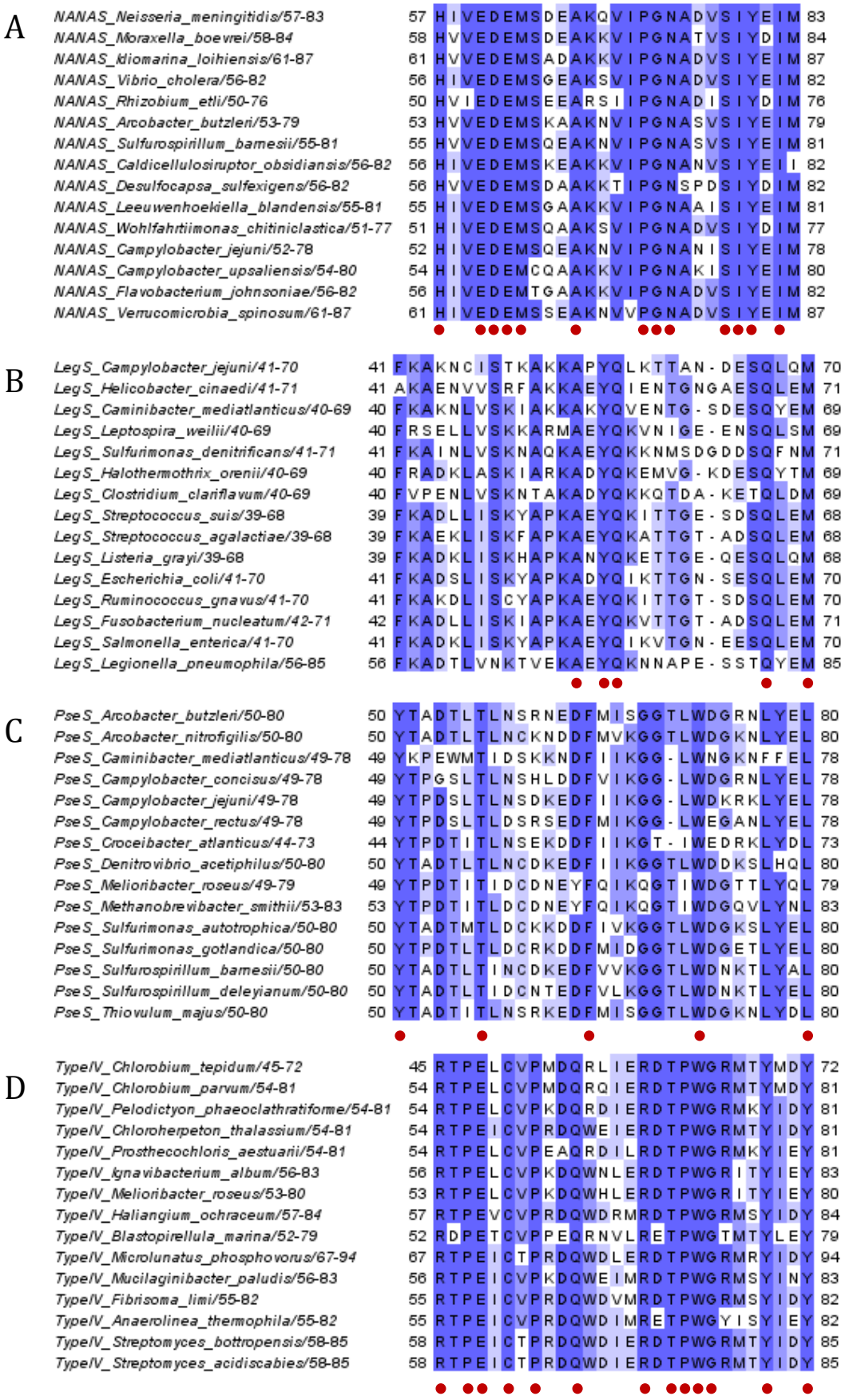


Figure 4.6 Alignment of SAS β2-α2 loop sequences displayed using Jalview (*159*). 15 sequences were selected from each of the four SAS clades. (A) 15 β2-α2 loop sequences from NANAS clade, (B) 15 β2-α2 loop sequences from LegS clade, (C) 15 β2-α2 loop sequences from PseS clade and (D) 15 β2-α2 loop sequences from Type IV clade. Residues highlighted in blue indicate conservation across the SAS family, with increasing intensity of blue indicating higher identity and absence of blue indicating lack of identity. Residues fully conserved within clade are marked with a red dot.

As depicted in Figure 4.6, the β2-α2 loops of both LegSs and PseSs are slightly longer than that of NANASs, which arises from the full SAS sequence alignment. The increased length may be to accommodate the slightly bulkier substrates of LegSs and PseSs which, as depicted in Figure 4.1, have an additional *N*-acetyl group when compared to ManNAc. The β2-α2 loop of Type IV SASs is also shorter than that of PseSs and LegSs and is a similar length to that for NANASs. The length of the Type IV β2-α2 loop may be indicative that the as yet unknown sugar substrate of this clade is a relatively simple member of the sialic acid family.

The second region of high intraclade conservation corresponds to the AFPL domain linker. The linker crosses over the catalytic site (Figure 4.7) and contributes a single phenylalanine residue into the active site of *Nme*NANAS. As aforementioned ^{Nme}Phe288 is only conserved in the NANAS clade. As shown in Figure 4.8, the residues comprising the AFPL domain linker differ significantly within the entire SAS family but show remarkably high conservation within each clade. Within this region there are several residues that are fully conserved within each of the four SAS clades (Figure 4.8). In the NANAS clade there is a highly conserved motif of ^{Nme}Asp285-Phe-Ala-Phe-Ala289 which contains residue ^{Nme}Phe288. In the LegS clade there is a highly conserved motif of Arg-Lys-Ser-Ile-Val-Ala and in the PseS clade a motif of Phe-Ala/Ser-Arg-Ser-Leu-Tyr in the corresponding locations.

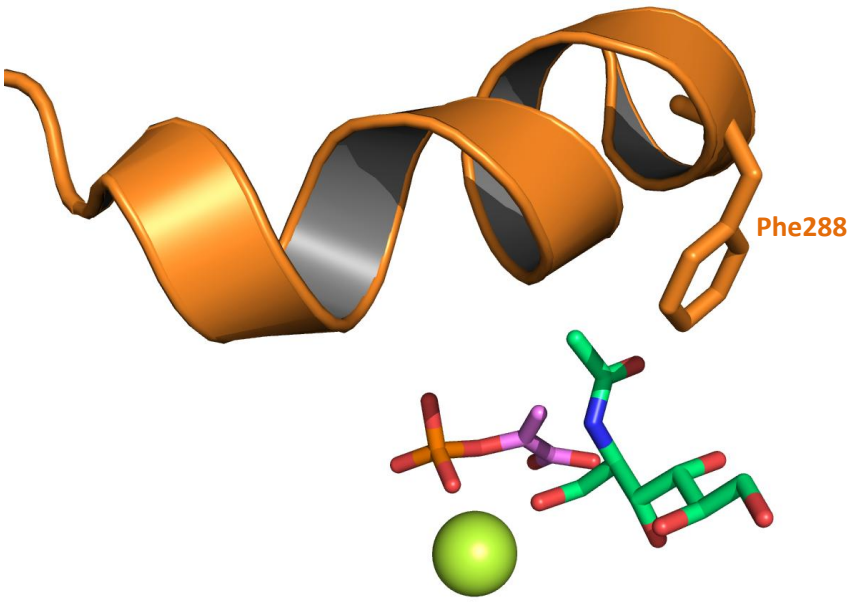


Figure 4.7 View of catalytic site and AFPL domain linker. AFPL domain linker is coloured in orange. ManNAc is coloured green, PEP pink and Mn^{2+} yellow. (PDB code (1XUZ) (*112*)).

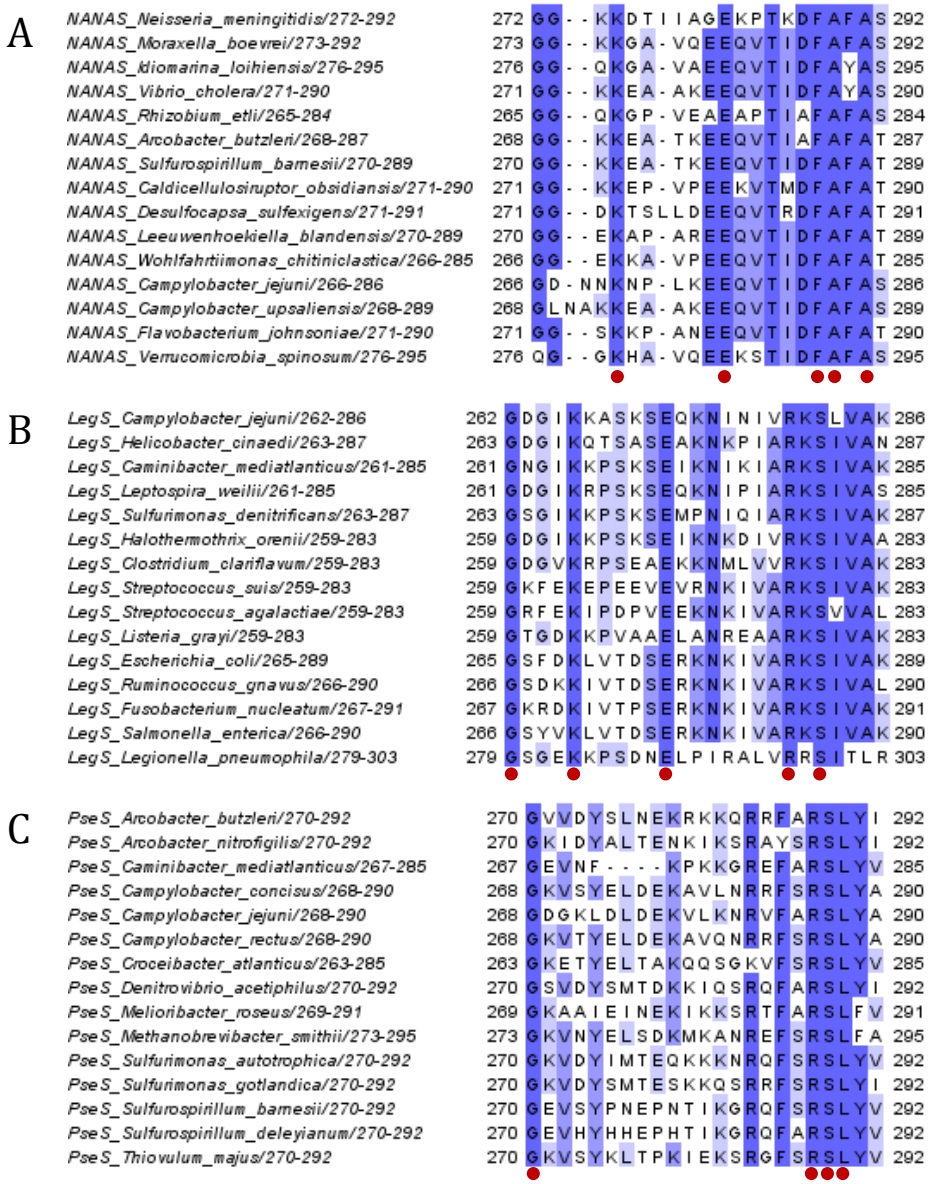


Figure 4.8 Alignment of SAS AFPL domain linker sequences displayed using Jalview (*159*). 15 sequences were selected from each of the four SAS clades. (A) 15 AFPL domain linker sequences from NANAS clade, (B) 15 AFPL domain linker sequences from LegS clade and (C) 15 AFPL domain linker sequences from PseS clade. Residues highlighted in blue indicate conservation across the SAS family, with increasing intensity of blue indicating higher identity and absence of blue indicating lack of identity. Residues fully conserved within clade are marked with a red dot.

4. 3 Cloning, expression and purification of *C. jejuni* SASs

4.3.1 Cloning of *C. jejuni* SASs

The genes encoding Cj1141 (NANAS), Cj1327 (LegS) and Cj1317 (PseS) from *C. jejuni* NCTC 11168 were amplified from genomic DNA using the nested PCR method outlined in Section 8.3.2. Gene-specific nested primers were designed incorporating generic extensions for further rounds of nested PCR (Table 8.8). Following successful amplification as determined by agarose gel electrophoresis, the product from the first round of nested PCR was purified. Purified first round product was extended using generic gateway primers, which incorporated sequence encoding an N-terminal TEV protease cleavage site (Table 8.8).

Successful amplification was determined by agarose gel electrophoresis and the product from the second round of nested PCR was also purified. Cloning and sub-cloning of the *C. jejuni* SAS genes were achieved through the use of the gateway system as outlined in Section 8.3. The linear product from second round nested-PCR was cloned into the donor vector pDONR-221. Ligated product was subsequently transformed into *E. coli* One Shot® TOP10 cells (Invitrogen) as outlined in Section 8.3.10. Plasmids from transformants were purified and each of the three *C. jejuni* SAS genes were successfully sequence verified in the donor vector. The three *C. jejuni* SAS genes were sub-cloned into the destination vector pDEST-17, which encodes an N-terminal His-tag.

4.3.2 Expression and purification

Each of the three *C. jejuni* SAS genes were transformed into three different expression cell lines to optimise the yield of soluble protein. The genes encoding *Cje*NANAS, *Cje*LegS and *Cje*PseS were each transformed into *E. coli* BL21 (DE3) Star cells, *E. coli* Rosetta™ 2 (DE3) cells and *E.coli* Bl21 (DE3) pBB540/pBB542 (Chaperone 3) cells. In order to compare induced against non-induced samples, two cultures of each cell line were grown overnight in LB, with one induced by IPTG and the other not. The best yield of soluble protein was attained using Chaperone 3 cells as analysed by SDS PAGE. *Cje*NANAS, *Cje*LegS and *Cje*PseS were thus expressed and purified from this cell line.

Each of the *C. jejuni* SASs were expressed and purified using a similar method to that used for *Nme*NANAS wild-type as outlined in Section 8.4. Cell pellets were resuspended in buffer and lysed via sonication. The resulting lysate was clarified by centrifugation to separate out the cell debris. Soluble protein was filtered and run through an IMAC column to purify the His-tagged protein. After the first round of IMAC purification, protein was desalted and treated with TEV protease to remove the His-tag and then run through a second IMAC column to separate the tag and TEV protease from the target protein. SEC was used as a final, polishing step and fractions were subsequently pooled, concentrated, flash-frozen and stored at -80 °C. Samples of purified enzyme were analysed by SDS PAGE (Figure 4.9).

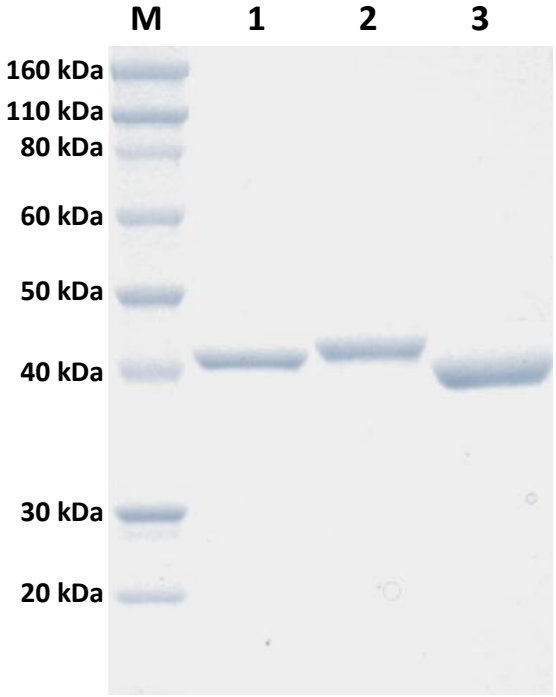


Figure 4.9 SDS-PAGE gel of purified *C. jejuni* SAS enzymes. Lane M contains marker with molecular weights defined. Lane 1 contains *Cje*NANAS (Cj1141), lane 2 *Cje*PseS (Cj1317) and lane 3 *Cje*LegS (Cj1327).

4.4 Kinetic analysis of *C. jejuni* SASs

4.4.1 Michaelis-Menten kinetics

To determine the kinetic parameters of the *C. jejuni* SAS enzymes with respect to NANA synthesis, the consumption of PEP was monitored using the assay described in Section 8.5.1. LegS and PseS activity were not tested for any of the *C. jejuni* SAS enzymes as the substrates for Leg and Pse biosynthesis (2,4-diacetamido-2,4,6-trideoxymannose and 2,4-diacetamido-2,4,6-trideoxaltrose respectively) were not commercially available and their chemical synthesis is complex (*161*).

Cuvettes containing 50 mM BTP (pH 8), 1 mM $MnCl_2$, PEP (varied) and ManNAc (varied) were incubated for 10 min at 25 °C. The enzymatic reaction was initiated by the addition of purified enzyme. Michaelis-Menten steady-state approximations were utilised to determine kinetic parameters (Figure 4.10). Kinetic parameters were compared to the values ascertained for wild-type *Nme*NANAS (Table 4.2).

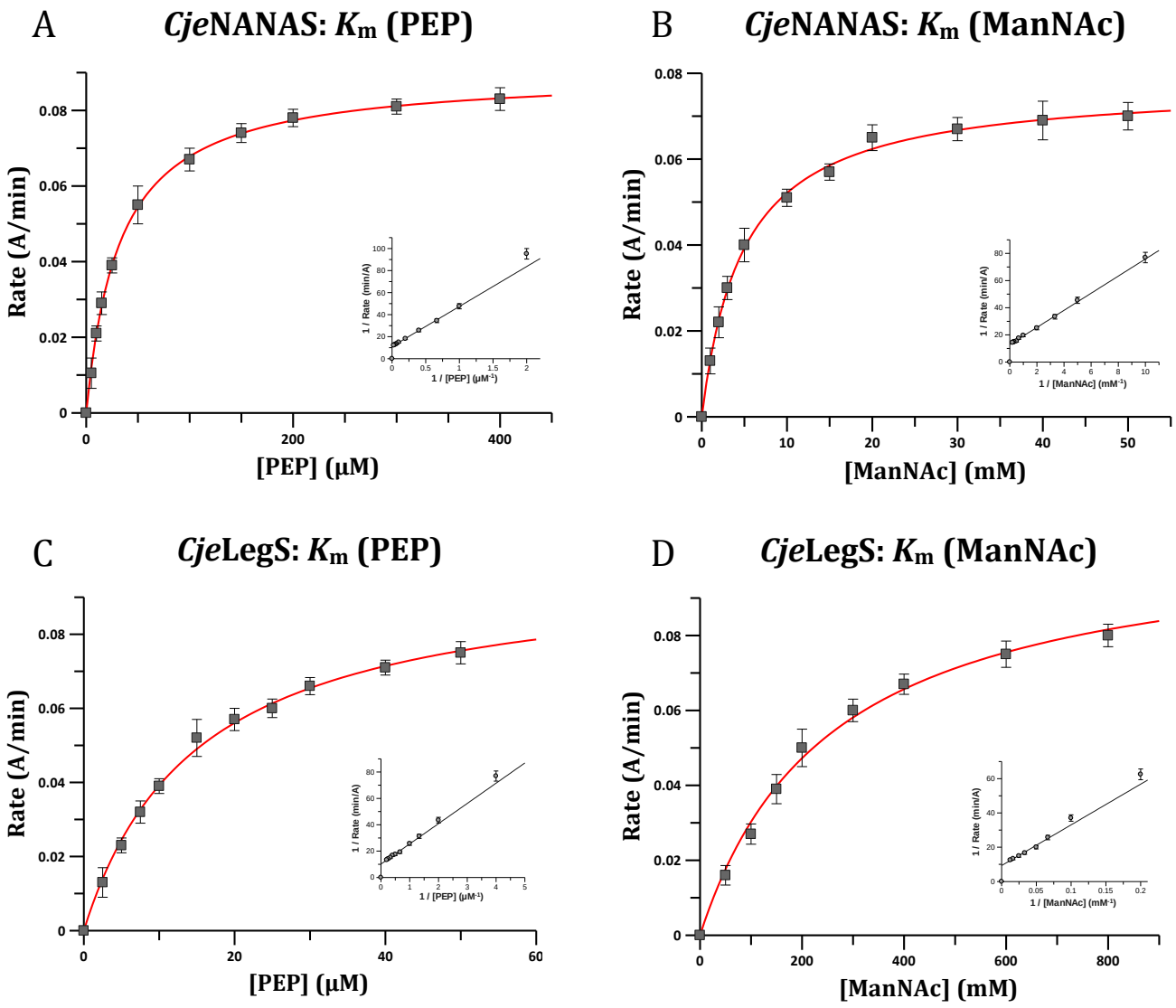


Figure 4.10 Michaelis-Menten curves for *Cje*NANAS and *Cje*LegS with ManNAc and PEP. (A) *Cje*NANAS PEP K_m curve, (B) *Cje*NANAS ManNAc K_m curve, (C) *Cje*LegS PEP K_m curve and (D) *Cje*LegS ManNAc K_m curve. Data points are shown in black and a line of best fit in red. A double reciprocal plot is displayed as an inset for each graph.

	*Nme*NANAS	*Cje*NANAS	*Cje*LegS	*Cje*PseS
k_{cat} (s^{-1})	3.1 ± 0.1	1.7 ± 0.1	0.05 ± 0.01	-
K_m PEP (μM)	28 ± 3	33 ± 1	15 ± 1	-
K_m ManNAc (mM)	2.9 ± 0.2	4.9 ± 0.2	256 ± 23	-
k_{cat}/K_m ManNAc (s^{-1}/M)	1068 ± 108	346 ± 35	0.21 ± 0.05	-

Table 4.2 Kinetic parameters for *Cje*NANAS, *Cje*LegS and *Cje*PseS compared with *Nme*NANAS. Values with - indicates k_{cat} of enzyme is less than 0.004 s^{-1}.

*Cje*NANAS was determined to have relatively similar kinetic parameters to *Nme*NANAS. The K_m (PEP) for *Cje*NANAS was almost the same as that calculated for *Nme*NANAS (33 ± 1 μM and 28 ± 3 μM respectively) whilst the K_m (ManNAc) for *Cje*NANAS was slightly larger than that for *Nme*NANAS (4.9 ± 0.2 mM and 2.9 ± 0.2 mM respectively). The k_{cat} for *Cje*NANAS was determined to be approximately half that for *Nme*NANAS at 1.7 ± 0.1 s^{-1} and 3.1 ± 0.1 s^{-1} respectively.

The kinetic values determined for *Cje*NANAS differ to those previously published in the literature by Sundaram *et al* (*90*). The K_m (PEP) and K_m (ManNAc) values formerly reported for *Cje*NANAS were 7.3 mM and 17.6 mM respectively with a k_{cat} of 0.3 s^{-1} (*90*). The differences in kinetic parameters attained may be attributed to several key experimental variations. The kinetic assay used in this study was a continuous assay which monitored the loss of PEP, whereas the assay used in the literature was a thiobarbituric acid stopped assay. Furthermore, Sundaram *et al* characterised *Cje*NANAS as an N-terminal His_6-S-tag chimera as opposed to an un-tagged enzyme. The incorporation of the N-terminal solubility extension may account for the larger K_m values and lower k_{cat} as it is possible that the solubility tag could interfere with the enzyme's ability to catalyse the NANAS reaction through preclusion of the active site. Therefore, the His_6-S-tagged enzyme may not accurately portray *in vivo* kinetic parameters for *Cje*NANAS.

Lastly, in this study *Cje*NANAS was assayed at 25 °C whereas the literature assay was done at 37 °C. 25 °C was chosen to keep consistency between each of the wild-type enzymes assayed in this thesis as many of the enzymes were found to be unstable in the absence of ligand (described in Section 4.5.2). It is possible that the higher temperature used to assay *Cje*NANAS in the literature had a negative effect on the enzyme's stability and thus the kinetic parameters as well.

Minimal NANAS activity has been previously reported for *Cje*LegS, however, kinetic parameters were not quantified (*64*). The K_m (PEP) and K_m (ManNAc) for *Cje*LegS were calculated to be 15 ± 1 μM and 256 ± 23 mM respectively. The K_m (PEP) for *Cje*LegS is approximately half that observed for either *Nme*NANAS of *Cje*NANAS. In comparison, the K_m (ManNAc) for *Cje*LegS is over 50 times greater than that observed for *Cje*NANAS and over 80 times greater than that observed for *Nme*NANAS. As previously mentioned, 2,4-diacetamido-2,4,6-trideoxymannose is not readily available, thus no kinetic parameters have been published in the literature for a LegS with the natural sugar substrate for Leg biosynthesis. As no K_m or k_{cat} values are available for *Cje*LegS with respect to 2,4-diacetamido-2 4,6-trideoxymannose, it is not possible to directly compare the catalytic efficiency of the enzyme between NANAS and LegS activity. However, it can be noted that *Cje*LegS is far less efficient than *Cje*NANAS at synthesising NANA.

*Cje*PseS was assayed for NANAS activity under the same physiological conditions previously used to determine PseS activity (*89*) as well as that described in Chapter 2 for *Nme*NANAS. The aforementioned single-step assay was used instead of the coupled assay described in the literature, and 2,4-diacetamido-2,4,6-trideoxyaltrose was substituted with ManNAc. Up to 35 μg of purified enzyme was used to initiate reaction (10 fold greater than that used to assay *Cje*PseS in the literature), however, no NANAS activity was observed. It was previously observed that Co^{2+} was more activating than Mn^{2+}, thus activity was also tested using Co^{2+}. Activity was tested with ManNAc at each step of the purification process, however, no loss of PEP was detected at any stage. This correlates with previous studies which concluded that *Cje*PseS is the only *C. jejuni* SAS without NANAS activity (*64*).

4.4.2 Attempted alternative substrate analysis

*Cje*NANAS has been previously kinetically characterised with several *N*-acetylmannosamines with modified acyl chains (*90*). The *N*-acetyl group of ManNAc was modified to a propionyl, butanoyl and pentanoyl and successfully reacted with PEP using *Cje*NANAS (*90*). Alternative sugar substrates have also been successfully observed with *S. agalacticae* NANAS (*105*). In order to probe the ability of the *C. jejuni* SASs to use alternative substrates, several ManNAc analogues were assayed (Table 4.3). However, no detectable activity could be measured for any of the three *C. jejuni* SASs with the ManNAc analogues depicted below, even in the presence of high enzyme concentrations (approximately 1 mg/mL) and high sugar concentration (up to 500 mM).

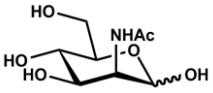
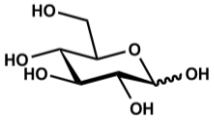
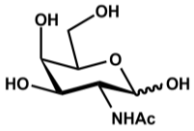
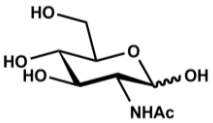
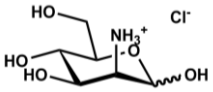
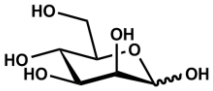
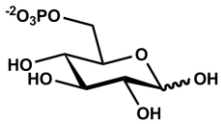
Name	Structure	Name	Structure
***N*-acetyl-D-mannosamine**		**D-glucose**	
***N*-acetyl-D-galactosamine**		**2-deoxy-D-glucose**	
***N*-acetyl-D-glucosamine**		**D-mannosamine**	
D-mannose		***N*-acetyl-D-mannosamine-6-phosphate**	

Table 4.3 Structures of ManNAc analogues assayed with *Cje*NANAS, *Cje*PseS and *Cje*LegS.

4.5 Binding studies of *C. jejuni* SASs

4.5.1 Isothermal titration calorimetry

Binding constants for Mn^{2+} and PEP were determined for each of the three *C. jejuni* SASs using the methods described in Section 8.5.4. ITC results from *Cje*NANAS, *Cje*PseS and *Cje*LegS were compared to previously determined dissociation constants (K_d) for *Nme*NANAS wild-type (Table 4.4) (*114*). Raw experimental and one-site model fitted ITC data is displayed in Figure 4.11.

	***Nme*NANAS (*114*)**	***Cje*NANAS**	***Cje*PseS**	***Cje*LegS**
K_d Mn^{2+} (μM)	10.6 ± 0.5	3.9 ± 0.1	4.0 ± 0.2	13 ± 1
K_d PEP (μM)	219 ± 11	320 ± 10	347 ± 9	356 ± 14

Table 4.4 ITC derived binding constants for wild-type *Nme*NANAS, *Cje*NANAS, *Cje*PseS and *Cje*LegS.

To determine the K_d value for metal with each of the *C. jejuni* SASs, Mn^{2+} was titrated into metal-free enzyme. The Mn^{2+} affinities of *Cje*NANAS and *Cje*PseS were determined to be approximately two-fold greater than that for *Nme*NANAS and three-fold greater than for *Cje*LegS. As depicted in Figure 4.11, the binding of Mn^{2+} is an exothermic event in each of the three *C. jejuni* enzymes and thus an enthalpically favourable process.

The K_d (PEP) values of *Cje*NANAS, *Cje*PseS and *Cje*LegS were determined to be very similar at 320 ± 10 μM, 347 ± 9 μM and 356 ± 14 μM respectively. The PEP affinities for each of the three *C. jejuni* SASs were slightly lower than that observed for *Nme*NANAS wild-type. As shown in Figure 4.11, the binding of PEP is an endothermic event in each of the three *C. jejuni* enzymes. This correlates with previous observations for PEP binding with wild-type *Nme*NANAS that was determined to be an entropically favourable event.

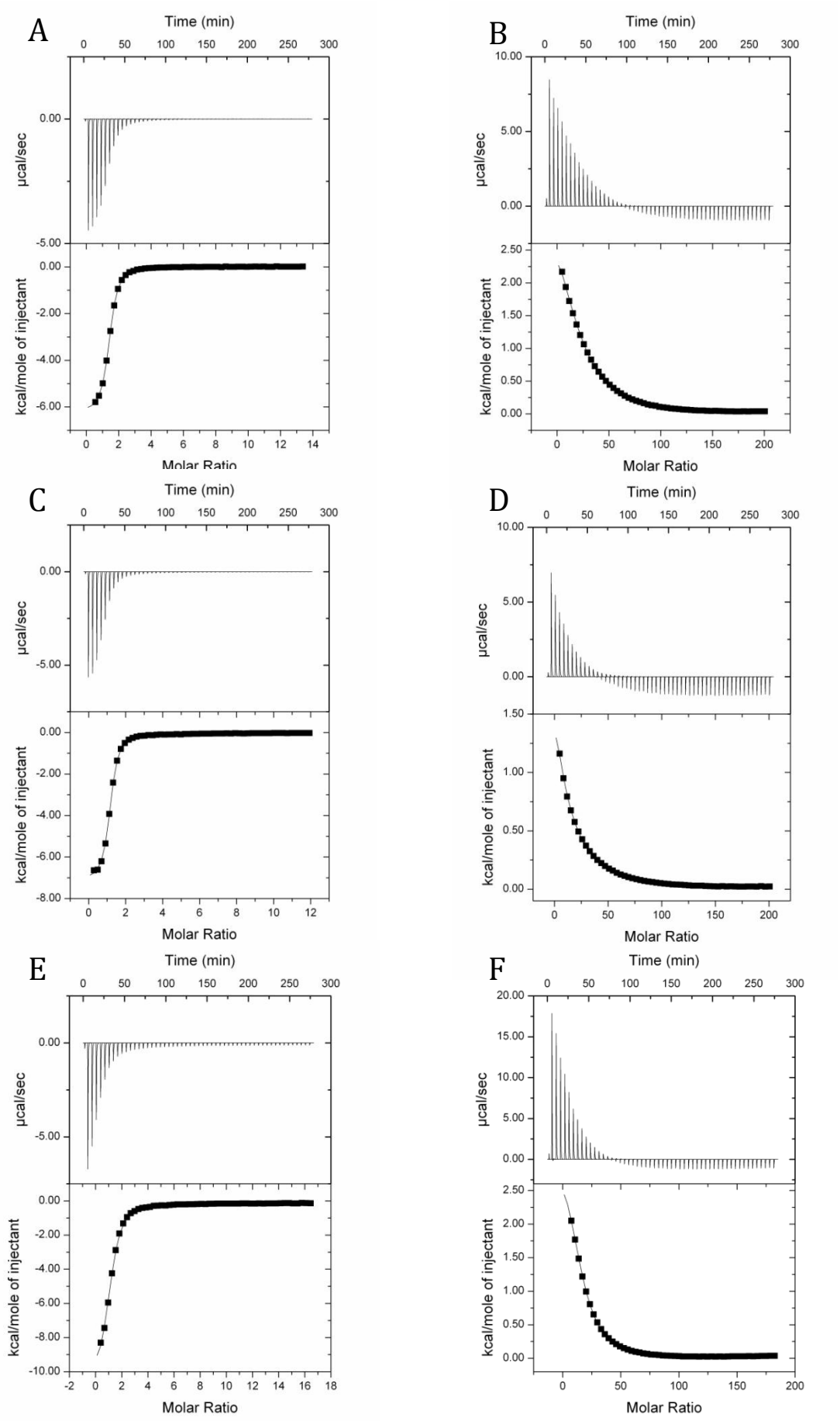


Figure 4.11 Enthalpy changes measured by isothermal titration calorimetry. (A) Mn^{2+} (5 mM) titrated into metal-free *Cje*NANAS (65 µM). (B) PEP (50 mM) titrated into *Cje*NANAS (40 µM) in the presence of Mn^{2+} (1 mM). (C) Mn^{2+} (5 mM) titrated into metal-free *Cje*PseS (65 µM). (D) PEP (50 mM) titrated into *Cje*PseS (40 µM) in the presence of Mn^{2+} (1 mM). (E) Mn^{2+} (5 mM) titrated into metal-free *Cje*LegS (65 µM). (F) PEP (50 mM) titrated into *Cje*LegS (40 µM) in the presence of Mn^{2+} (1 mM).

4.5.2 Thermal stability

The effects of Mn^{2+}, PEP and ManNAc/rManNAc on the thermal stability of the three *C. jejuni* SASs were assessed by DSF using the method outlined in Section 8.5.3. Enzymes used for DSF were treated with 10 mM EDTA prior to SEC to ensure removal of trace metals. The melting points determined for each sample were plotted and compared with data obtained for *Nme*NANAS wild-type (Figure 4.12 and Table 4.5).

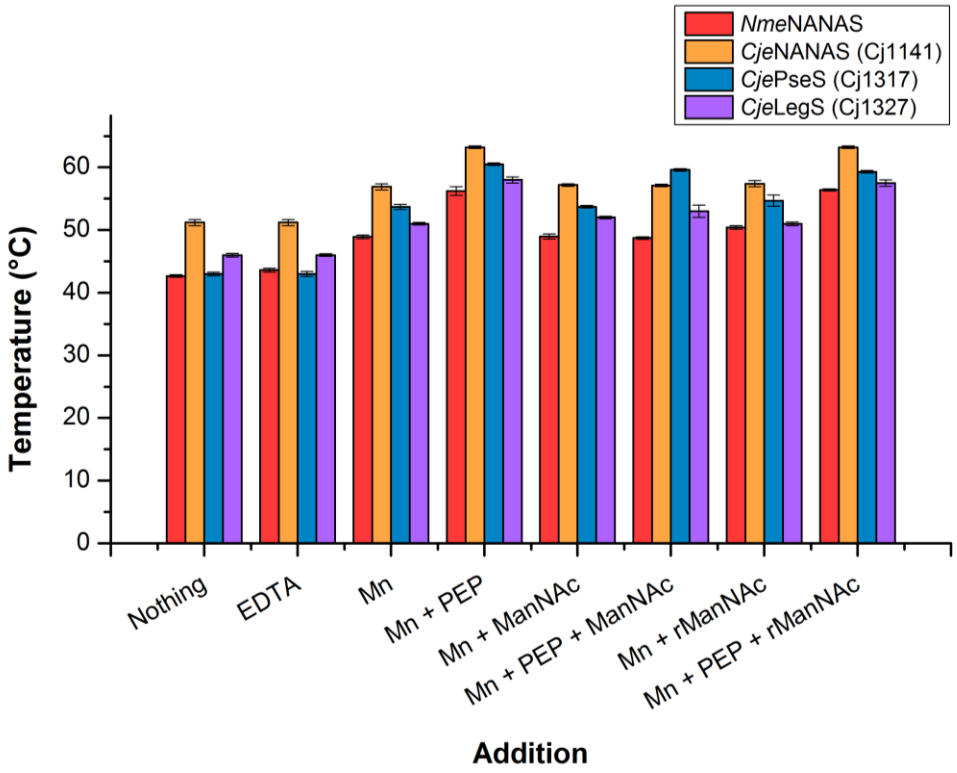


Figure 4.12 Bar graph of thermal melting points for *Nme*NANAS (red), *Cje*NANAS (orange), *Cje*PseS (blue) and *Cje*LegS (purple) in the presence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc at 10 mM. ManNAc/rManNAc was used at 1 mM when in combination with PEP.

Addition	*Nme*NANAS (°C)	*Cje*NANAS (°C)	*Cje*PseS (°C)	*Cje*LegS (°C)
Nothing	42.7 ± 0.2	51.2 ± 0.5	43.0 ± 0.3	45.8 ± 0.3
EDTA	43.6 ± 0.3	51.2 ± 0.5	42.8 ± 0.4	46.0 ± 0.2
Mn^{2+}	48.9 ± 0.3	56.9 ± 0.5	53.7 ± 0.4	50.9 ± 0.2
Mn^{2+} + PEP	56.2 ± 0.7	63.2 ± 0.2	60.5 ± 0.2	58.3 ± 0.5
Mn^{2+} + ManNAc	49.0 ± 0.4	57.2 ± 0.2	53.7 ± 0.2	52.0 ± 0.2
Mn^{2+} + PEP + ManNAc	48.7 ± 0.2	57.1 ± 0.2	59.6 ± 0.2	53.1 ± 0.5
Mn^{2+} + rManNAc	50.4 ± 0.3	57.4 ± 0.5	54.7 ± 0.9	51.1 ± 0.3
Mn^{2+} + PEP + rManNAc	56.4 ± 0.2	63.2 ± 0.2	59.3 ± 0.2	57.5 ± 0.5

Table 4.5 Thermal melting points for *Nme*NANAS, *Cje*NANAS, *Cje*PseS and *Cje*LegS in the presence and absence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc at 10 mM. ManNAc/rManNAc was used at 1 mM when in combination with PEP.

Apoenzyme *Cje*NANAS, *Cje*PseS and *Cje*LegS melted at 51.2 ± 0.5 °C, 43.0 ± 0.3 °C and 45.8 ± 0.3 °C respectively. Shown in Figure 4.12, *Cje*NANAS was found to have slightly higher thermal stability than *Cje*PseS, *Cje*LegS and *Nme*NANAS under each of the conditions tested. As shown in Figure 4.12, the addition of 1 mM EDTA to apoenzyme did not influence the thermal stability of *Cje*NANAS, *Cje*PseS or *Cje*LegS. The absence of an effect caused by EDTA is likely due to the successful removal of trace metal ions during purification. The addition of 1 mM Mn^{2+} increased the melting points of *Cje*NANAS, *Cje*PseS and *Cje*LegS by approximately 6 °C, 11 °C and 5 °C respectively. This increased thermal stability in the presence of Mn^{2+} is also observed in *Nme*NANAS.

The largest thermal shift observed for each of the three *C. jejuni* enzymes was in the presence of both Mn^{2+} and PEP. The combined addition of 1 mM Mn^{2+} and 1 mM PEP to *Cje*NANAS, *Cje*PseS and *Cje*LegS increased the melting point of apoenzyme by approximately 12 °C, 17.5 °C and 12 °C respectively. The observation that Mn^{2+} and PEP are so significantly stabilising has been previously noted for *Nme*NANAS (Chapter 3). For each of the *C. jejuni* SASs, the addition of 10 mM ManNAc or rManNAc in combination with 1 mM Mn^{2+} did not have any additional effect to that observed for only Mn^{2+}.

The melting temperatures for both *Cje*NANAS and *Cje*LegS in the presence of 1 mM Mn^{2+}, 1 mM PEP and 1 mM ManNAc were similar to that observed for enzyme in the presence of Mn^{2+}. As described in Section 4.4.2, both *Cje*NANAS and *Cje*LegS are able to synthesise NANA. Therefore, the melting temperatures observed for *Cje*NANAS and *Cje*LegS in the presence of Mn^{2+}, PEP and ManNAc are likely caused by the presence of Mn^{2+} and NANA. It is unlikely that NANA would favourably bind to either *Cje*NANAS or *Cje*LegS given that it is the product of the reaction catalysed and would need to be removed from the active site for further reactions to progress. The resulting thermal stability observed for *Cje*NANAS and *Cje*LegS in the presence of Mn^{2+}, PEP and ManNAc may consequentially be contributed solely by Mn^{2+}, hence the similar melting temperatures observed when compared to enzyme in the presence of only Mn^{2+}.

Conversely, *Cje*PseS in the presence of Mn^{2+}, PEP and ManNAc has a similar melting temperature to *Cje*PseS in the presence of Mn^{2+} and PEP. As described in Sections 4.3.1 and 4.4.1, *Cje*PseS is able to bind PEP but does not catalyse the formation of NANAS. Given that the presence of ManNAc has little to no effect on the thermal stability of any of the SASs tested, it is likely that the thermal shift observed for *Cje*PseS in the presence of Mn^{2+}, PEP and ManNAc is caused by PEP still being present in the active site. In the presence of Mn^{2+}, PEP and rManNAc, each of the three *C. jejuni* SASs have similar melting temperatures to enzyme in the presence of Mn^{2+} and PEP. This is likely caused by an inability to turn over rManNAc and the retained presence of PEP in the active site for each enzyme.

PEP binding in wild-type *Nme*NANAS has been shown to trigger a hydrogen bonding network that rearranges the active site for ManNAc binding (Chapter 3). Given the similar observations for both ITC and DSF with respect to PEP between the *C. jejuni* SASs and *Nme*NANAS, it is highly likely that an analogous mechanism is employed by each of these enzymes. Furthermore, residue ^{Nme}Glu134, which was identified to be involved in the hydrogen bonding network is highly conserved in NANAS, LegS and PseS enzymes.

4.6 Structural analysis of *C. jejuni* SASs

4.6.1 Circular dichroism

CD was used to assess and compare the secondary structures of *Cje*NANAS, *Cje*PseS and *Cje*LegS. CD spectra of the *C. jejuni* SASs are compared with *Nme*NANAS data in Figure 4.13. The K2D3 server (*142*) was used to estimate the secondary structure composition from CD spectra. Shown in Table 4.6, estimated secondary structure is compared to values derived from the crystal structure of *Nme*NANAS (PDB code: 1XUZ) and computational models of each of the *C. jejuni* enzymes (models are described in Section 4.6.6).

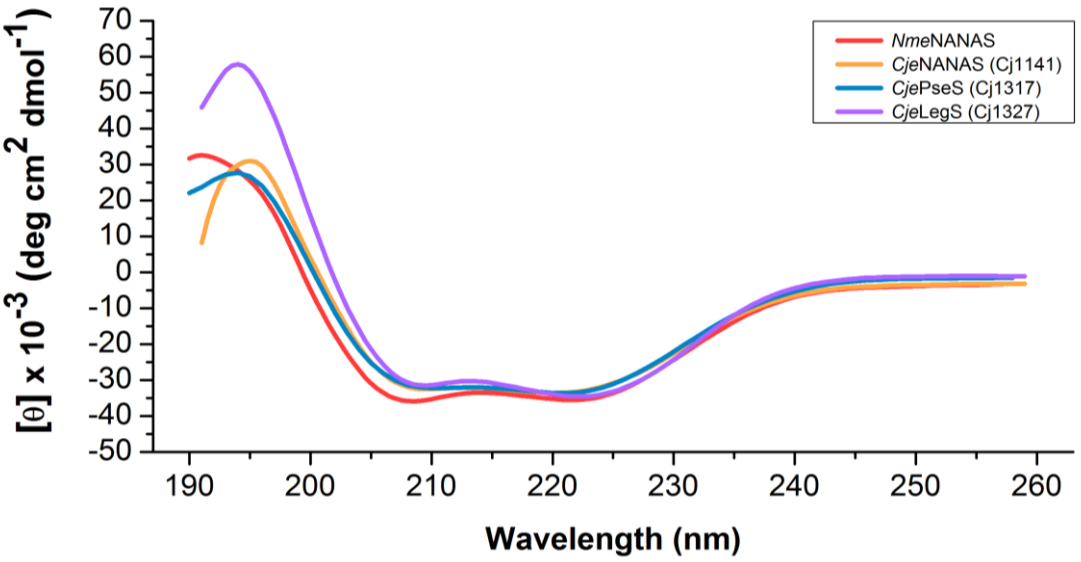


Figure 4.13 CD spectra for *Nme*NANAS (red), *Cje*NANAS (orange), *Cje*PseS (blue) and *Cje*LegS (purple).

	*Nme*NANAS	*Nme*NANAS (structure)	*Cje*NANAS	*Cje*NANAS (Model)	*Cje*PseS	*Cje*PseS (Model)	*Cje*LegS	*Cje*LegS (Model)
α-helix	59.9 %	50.3 %	51.3 %	48.1 %	54.3 %	54.1 %	54.2 %	50.0 %
β-sheet	12.3 %	15.5 %	13.6 %	16.5 %	13.3 %	15.2 %	13.3 %	16.4 %
other	27.8 %	34.2 %	35.1 %	35.4 %	32.4 %	30.7 %	32.5 %	33.6 %

Table 4.6 Predictions of secondary structure composition for *Nme*NANAS, *Cje*NANAS, *Cje*PseS and *Cje*LegS calculated from CD data using K2D3 server (*142*). Secondary structure composition of *Nme*NANAS structure calculated from crystallographic coordinates 1XUZ.

As depicted in Figure 4.13 and Table 4.6, each of the *C. jejuni* enzymes are folded in solution and have comparable secondary structure compositions. The secondary structure predictions from K2D3 are similar to those calculated from the computational models of the *C.jejuni* enzymes. Furthermore, the secondary structure compositions of each of the three *C. jejuni* enzymes are relatively similar to *Nme*NANAS and indicates that these enzymes are all structurally analogous.

4.6.2 Mass spectrometry

The masses of each of the *C. jejuni* SASs were analysed using MS (Table 4.7). Each of the experimental masses obtained were almost identical to those calculated from the respective peptide sequences using ProtParam (*141*).

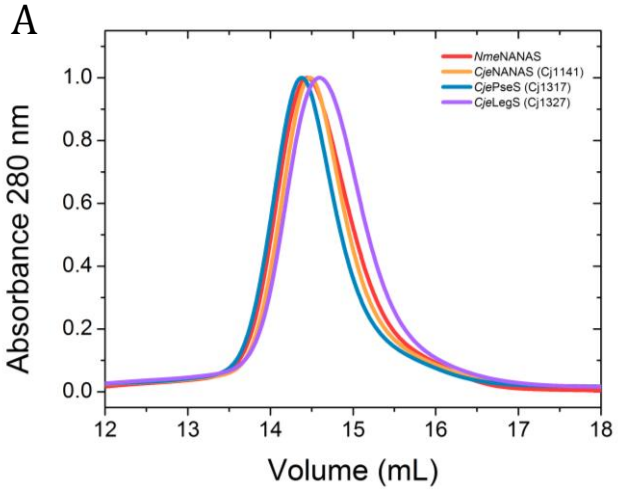
	***Cje*NANAS**	***Cje*PseS**	***Cje*LegS**
Experimental mass (Da)	42,111	42,275	40,719
Theoretical mass (Da)	42,111	42,276	40,720

Table 4.7 Mass spectrometry measurements of His_6-tagged *Cje*NANAS, *Cje*PseS and *Cje*LegS.

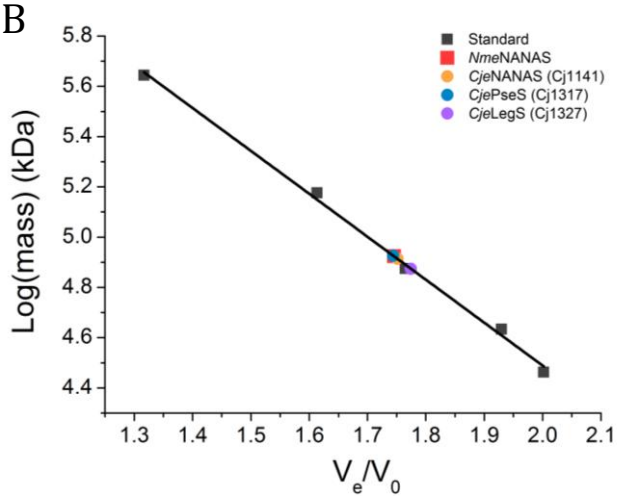
4.6.3 Analytical size exclusion chromatography

The quaternary structure of *Cje*NANAS, *Cje*PseS and *Cje*LegS was investigated using analytical SEC. The method used for analytical SEC is outlined in Section 8.5.6. A calibration curve was generated using protein standards of known molecular weight. The elution volume (V_e) of protein standards relative to the column void volume (V_0) was plotted against the log of the protein standards molecular mass. A linear fit of the analytical SEC calibration data was determined to have an R^2 value of 0.987. The calibration curve was used to extrapolate the molecular masses of each of the three *C. jejuni* SASs as shown in Figure 4.14.

A



B



C

	***Nme*NANAS**	***Cje*NANAS**	***Cje*PseS**	***Cje*LegS**
V_e/V_0	1.74	1.75	1.74	1.77
Estimated mass (kDa)	81	79	81	72
Theoretical mass of dimer (kDa)	77	77	77	74

Figure 4.14 (A) Normalised analytical SEC chromatograms for *Nme*NANAS (red), *Cje*NANAS (orange), *Cje*PseS (blue) and *Cje*LegS (purple). (B) Analytical SEC calibration curve. Protein standards are marked in grey, *Nme*NANAS in red, *Cje*NANAS in orange, *Cje*PseS in blue and *Cje*LegS in purple. A line of best fit is plotted in black. (C) Table of estimated and theoretical masses for *Nme*NANAS and *C. jejuni* SAS enzymes.

As shown in Figure 4.14, each of the three *C. jejuni* enzymes elute as a single species at molecular weights correpsonding to that calculated for dimeric protein. *Nme*NANAS has been previously identified as a dimer, and given the high sequence similarity of each of the *C. jejuni* SASs with *Nme*NANAS, it is unsurprising that they are dimeric as well. Shown in Figure 4.14A, *Nme*NANAS, *Cje*NANAS and *Cje*PseS elute at almost identical volumes, whilst *Cje*LegS elutes slightly later. This corresponds with the molecular masses of *Nme*NANAS, *Cje*NANAS and *Cje*PseS being very similar (approximately 77 kDa dimers) whilst the molecular mass of *Cje*LegS is slightly smaller (approximately 74 kDa dimer).

4.6.4 Small-angle X-ray scattering

SAXS profiles were determined to assess the solution structures of *Cje*NANAS, *Cje*PseS and *Cje*LegS. The methods used to determine the SAXS profiles of these enzymes are outlined in Section 8.5.8 and 8.5.9. The SAXS profiles of ligand-free *C. jejuni* SASs are displayed in Figure 4.15 and are overlaid with the ligand-free scattering profile of *Nme*NANAS.

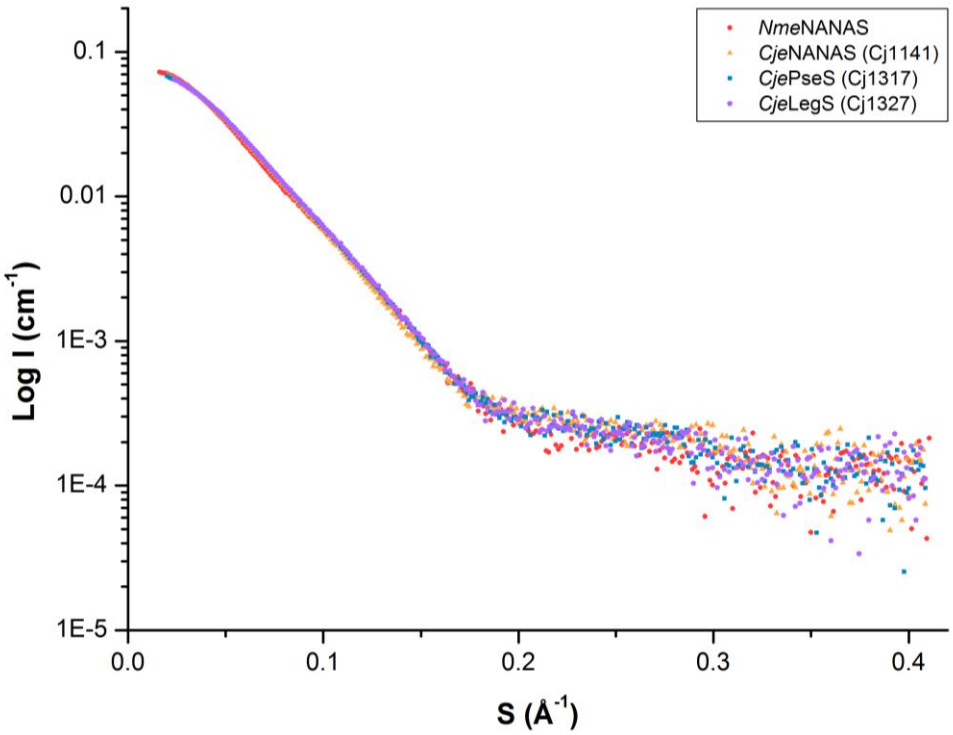


Figure 4.15 SAXS profiles of ligand-free *Nme*NANAS (red), *Cje*NANAS (orange), *Cje*PseS (blue) and *Cje*LegS (purple). Scattering data was collected using approximately 10 mg mL^{-1} protein in 30 mM triethanolamine-HCl buffer at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

As portrayed in Figure 4.15, each of the three *C. jejuni* SASs have remarkably similar SAXS profiles. Furthermore, the profiles are analogous to what is observed for ligand-free *Nme*NANAS. The structure parameters obtained from SAXS are also similar between each of the three *C. jejuni* enzymes (Table 4.8). The molecular masses of each of the *C. jejuni* SASs were calculated from SAXS data using SAXS MoW (*154*). The estimated masses for *Cje*NANAS, *Cje*PseS and *Cje*LegS were calculated to be 88, 84 and 83 kDa respectively, which is comparable to the 85 kDa mass estimated for *Nme*NANAS. The masses of the *C. jejuni* SASs indicate that the enzymes are dimeric under the conditions used for SAXS, which correlates with the results obtained from analytical SEC outlined in Section 4.6.3.

	*Nme*NANAS	*Cje*NANAS	*Cje*PseS	*Cje*LegS
I_0 (cm^{-1})	0.082 ± 0.001	0.079 ± 0.001	0.078 ± 0.001	0.078 ± 0.001
R_g (Å)	33.0 ± 0.5	32.3 ± 0.1	31.7 ± 0.2	31.8 ± 0.2
D_{max} (Å)	116	113	111	111
Porod volume (Å^3)	121,724	122,906	115,453	112,927
Estimated mass (kDa)	85	88	84	83
Theoretical mass of dimer (kDa)	77	77	77	74

Table 4.8 SAXS structural parameters determined for *C. jejuni* SASs compared with parameters for *Nme*NANAS. SAXS data obtained at a wavelength of 1.0332 Å.

Theoretical scattering curves were generated from atomic coordinates of the *C. jejuni* SAS models described in Section 4.6.6 and compared with experimental scattering curves using CRYSOL (*153*) (Figure 4.16). Each of the *C. jejuni* SAS models were generated from the liganded *Nme*NANAS 1XUZ template, yet none of the theoretical plots generated from the *C. jejuni* SAS models fit well with the respective experimental scattering profiles. As displayed in Figure 4.16A-C, each of the fits appear to be good at S values below 0.12 Å^{-1} but subsequently deteriorate in quality. The Chi values for *Cje*NANAS, *Cje*PseS and *Cje*LegS with their respective theoretical scattering profiles are 5.4, 5.4 and 5.8 respectively.

The poor fit between experimental SAXS profiles of ligand-free enzyme and theoretical scattering has been previously observed with *Nme*NANAS (Chapter 3). Although the ligand-free fit is poor, the experimental SAXS profile of *Nme*NANAS is altered in the presence of 1 mM Mn^{2+} and 1 mM PEP and the chi value of the fit is improved from 6.1 to 2.0. In Chapter 3, the conformational change observed with PEP-bound *Nme*NANAS has been attributed to PEP binding, reorganising the active site in preparation for ManNAc binding. As described in Section 4.5, the DSF and ITC profiles of each of the three *C. jejuni* enzymes with regards to PEP are very similar to what is observed with *Nme*NANAS. Given the similarities observed, it is highly likely that each of the *C. jejuni* SASs also undergoes a conformational change upon PEP binding and that the poor fits observed in Figure 4.16A-C are due to the absence of ligand.

Shown in Figure 4.16D, the theoretical scattering profiles from the models of *Cje*NANAS and *Cje*PseS align well with the profile generated from the crystallographic coordinates of *Nme*NANAS (1XUZ). The profile generated from the *Cje*LegS model also aligns well but is slightly different from the other three at higher S values.

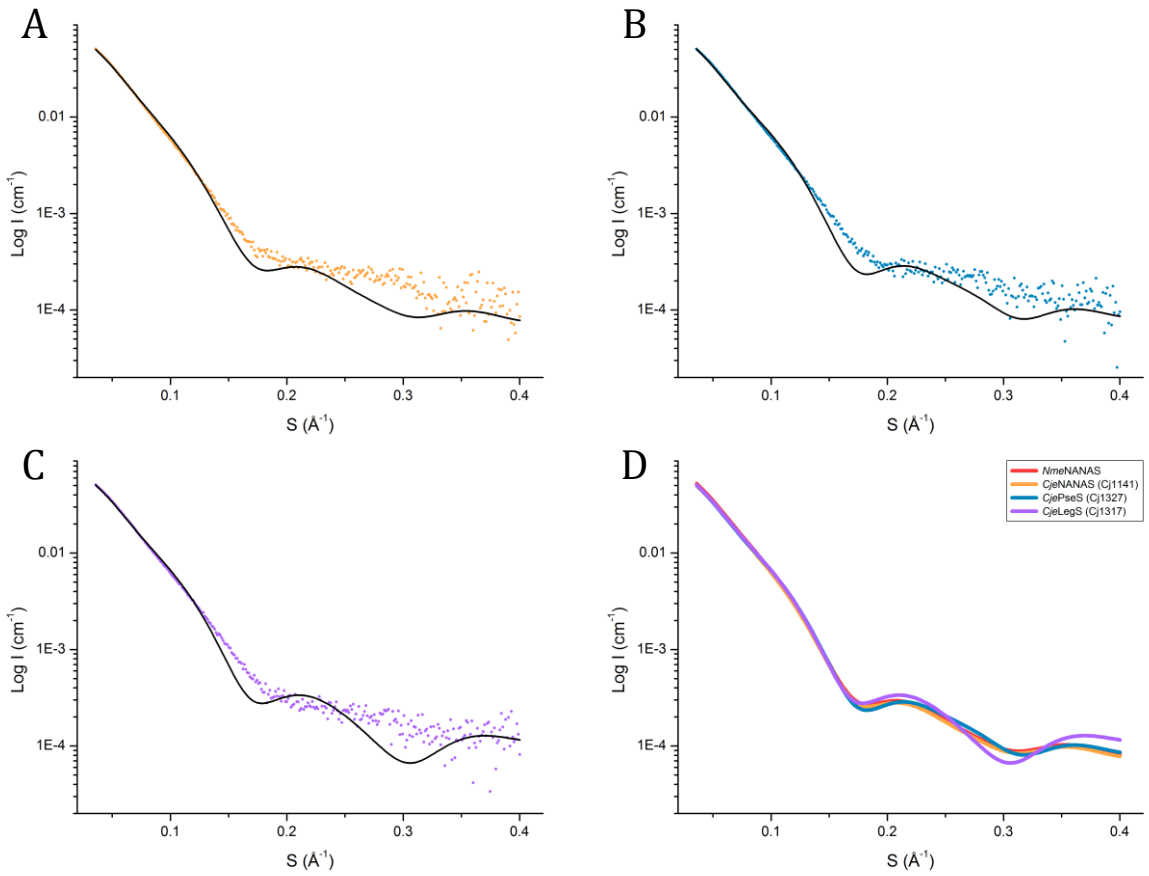


Figure 4.16 Comparisons of experimental and theoretical SAXS profiles using CRYSOL. (A) Comparison of SAXS profiles for ligand-free *Cje*NANAS (orange) with theoretical scattering generated from dimer model of *Cje*NANAS (black line). (B) Comparison of SAXS profiles for ligand-free *Cje*PseS (blue) with theoretical scattering generated from dimer model of *Cje*PseS (black line). (C) Comparison of SAXS profiles for ligand-free *Cje*LegS (purple) with theoretical scattering generated from dimer model of *Cje*LegS (black line). (D) Comparison of theoretical scattering profiles of *Nme*NANAS (red), *Cje*NANAS (orange), *Cje*PseS (blue) and *Cje*LegS (purple). Theoretical scattering for *Nme*NANAS was generated from crystallographic coordinates 1XUZ. SAXS data was collected using approximately 10 mg mL^{-1} protein in 30 mM triethanolamine-HCl buffer at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

4.6.5 Attempted crystallisation

Several attempts at crystallising the *C. jejuni* SASs were made. Each of the enzymes were screened using conditions modified from those previously described for crystallising *Nme*NANAS (*112*). Protein crystallisation conditions were also screened using commercially available Morpheus, PACT and JCSG screens. Protein was concentrated to 6 mg/mL and 15 mg/mL and screened in the absence and presence of ligands. Screens were incubated at 5 °C and 20 °C for several months, but none of the conditions that were screened yielded crystals of diffraction quality.

4.6.6 Modelling of *C. jejuni* SASs

Given that crystallisation of the *C. jejuni* SASs was unsuccessful, modelling was used to elucidate the structures of *Cje*NANAS, *Cje*PseS and *Cje*LegS. Models of *Cje*NANAS, *Cje*LegS and *Cje*PseS were generated from the SWISS-MODEL server (162) using the methods outlined in Section 8.8.3. The crystal structure of *Nme*NANAS (PDB code 1XUZ) was used as a template for each of the *C. jejuni* enzymes. Displayed in Figures 4.17-4.19 are the results from modelling of *Cje*NANAS, *Cje*PseS and *Cje*LegS. The C. jejuni SAS models were also aligned with the crystal structure of NmeNANAS for comparison.

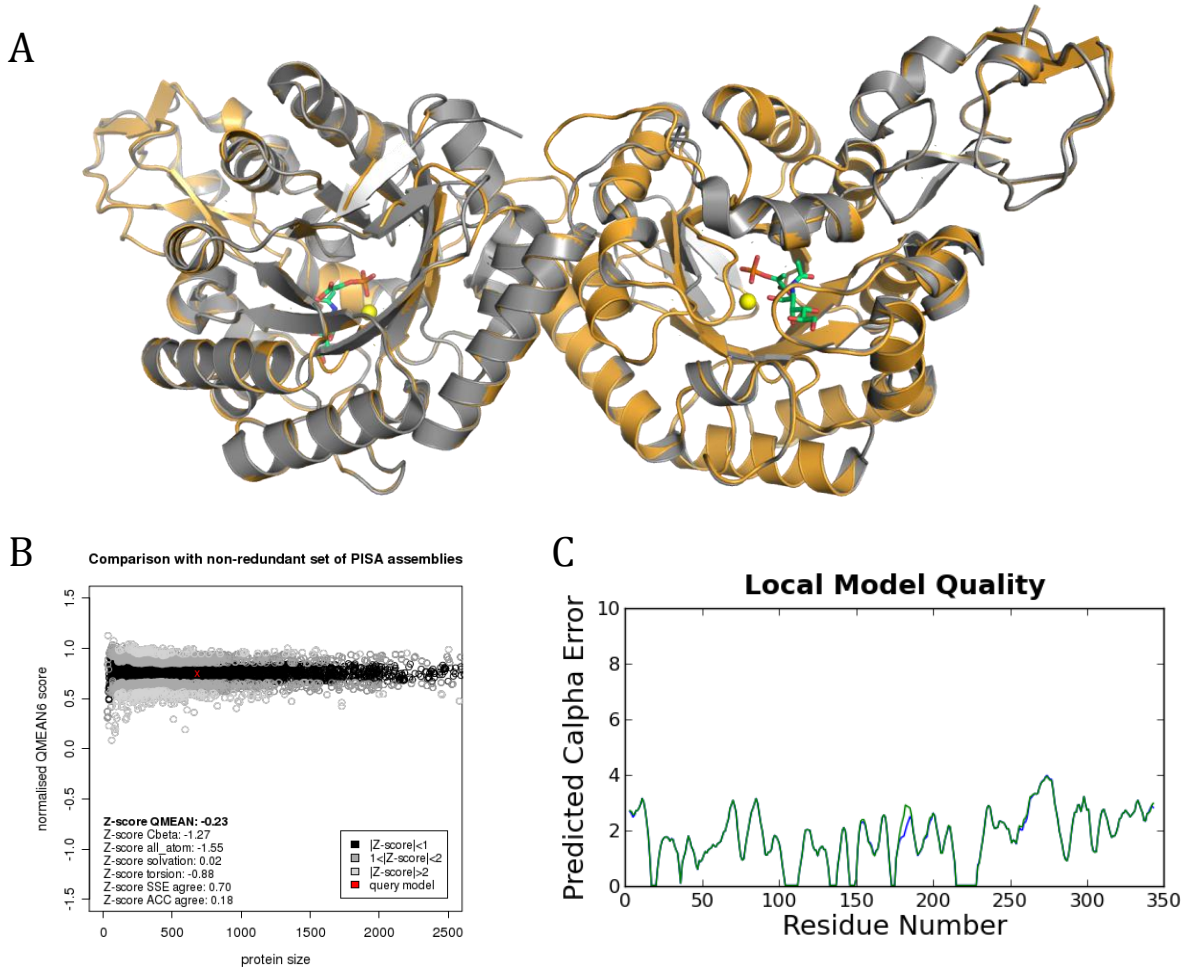


Figure 4.17 Modelling results of *Cje*NANAS generated using the SWISS-MODEL server. (A) *Cje*NANAS dimer aligned to *Nme*NANAS dimer (PDB code 1XUZ). *Cje*NANAS is coloured orange and *Nme*NANAS is coloured grey. ManNAc and PEP from *Nme*NANAS 1XUZ are coloured green and Mn^{2+} yellow. (B) Comparison of model QMEAN score with non-redundant PISA assemblies. (C) Graphical representation of local model quality. Chain A in green and chain B in blue.

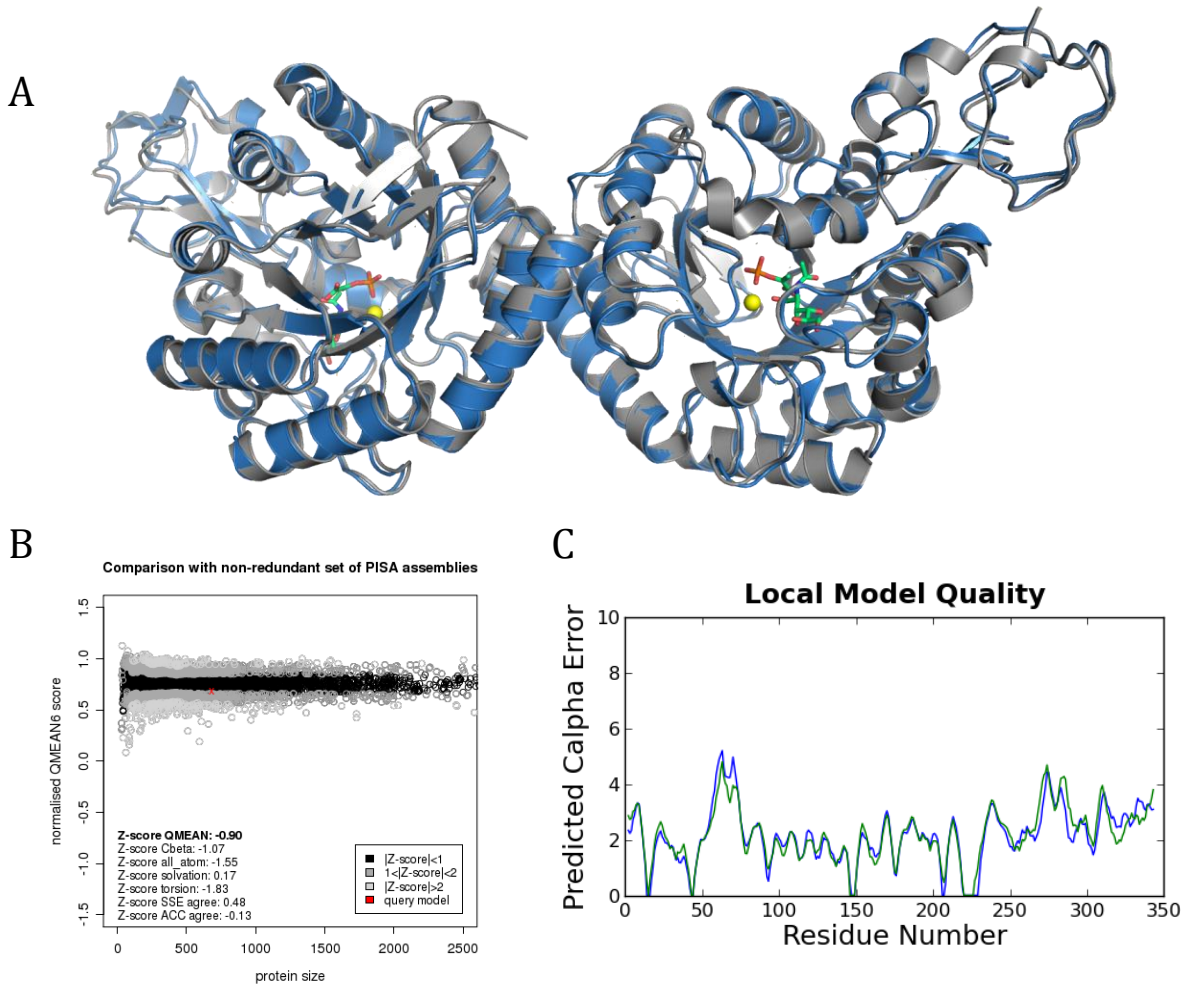


Figure 4.18 Modelling results of *Cje*PseS generated using the SWISS-MODEL server. (A) *Cje*PseS dimer aligned to *Nme*NANAS dimer (PDB code 1XUZ). *Cje*PseS is coloured blue and *Nme*NANAS is coloured grey. ManNAc and PEP from *Nme*NANAS 1XUZ are coloured green and Mn^{2+} yellow. (B) Comparison of model QMEAN score with non-redundant PISA assemblies. (C) Graphical representation of local model quality. Chain A in green and chain B in blue.

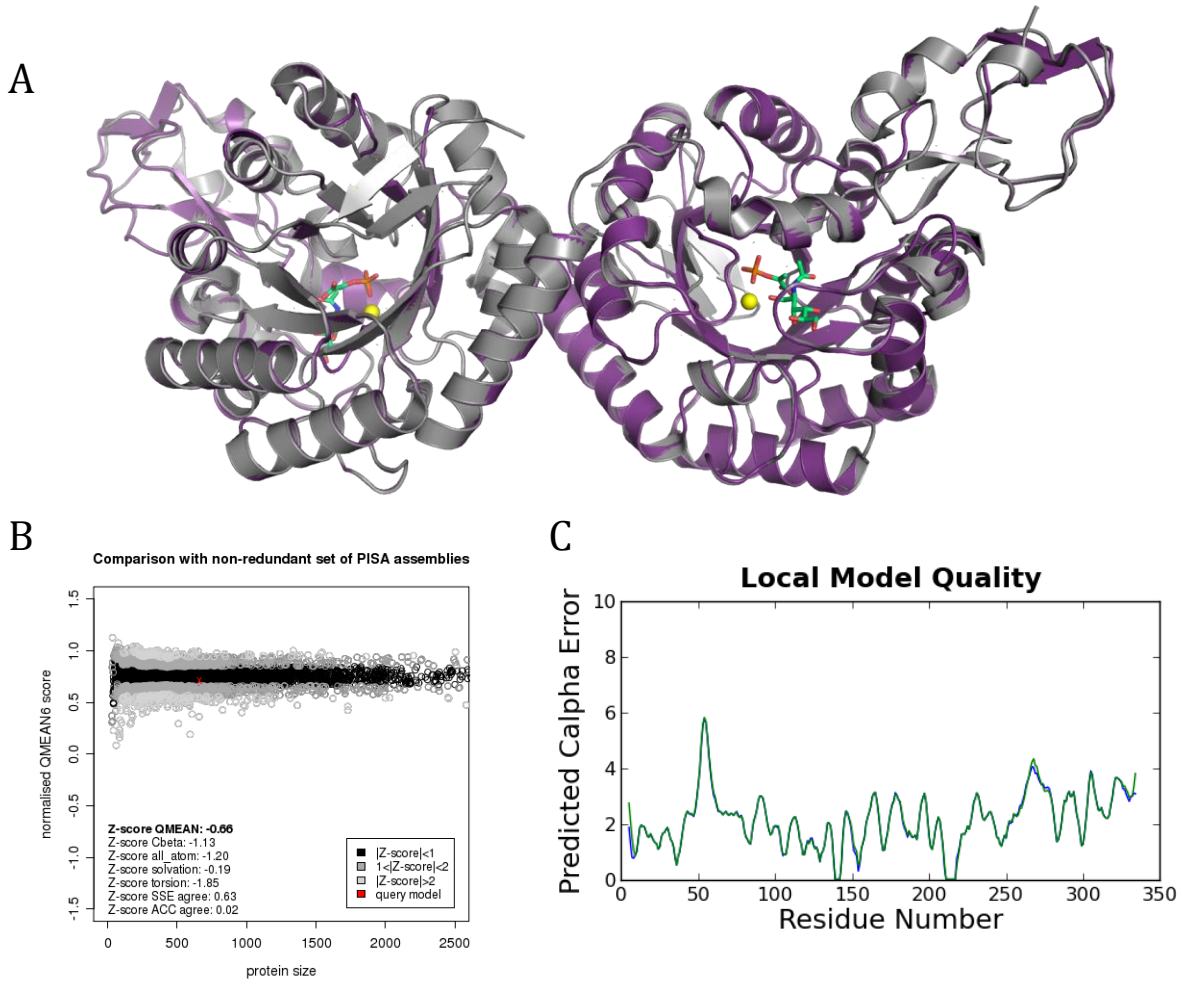


Figure 4.19 Modelling results of *Cje*LegS generated using the SWISS-MODEL server. (A) *Cje*LegS dimer aligned to *Nme*NANAS dimer (PDB code 1XUZ). *Cje*LegS is coloured purple and *Nme*NANAS is coloured grey. ManNAc and PEP from *Nme*NANAS 1XUZ are coloured green and Mn^{2+} yellow. (B) Comparison of model QMEAN score with non-redundant PISA assemblies. (C) Graphical representation of local model quality. Chain A in green and chain B in blue.

As expected, the overall folds of the catalytic and AFPL domains are maintained among each of the three models with comparison to the template structure of *Nme*NANAS 1XUZ (Figures 4.17-4.19). The root mean square deviation (RMSD) values for the models of *Cje*NANAS, *Cje*PseS and *Cje*LegS aligned to the *Nme*NANAS crystal structure are 0.25 Å (642 atoms to 642 atoms), 0.60 Å (619 atoms to 619 atoms) and 0.26 Å (530 atoms to 530 atoms) respectively.

The quality of the models were assessed using the quality model energy analysis (QMEAN) server in SWISS-MODEL (*163*). QMEAN is a composite scoring function utilising six different structural descriptors to assess the major geometrical aspects of a protein model. Models with QMEAN scores closest to 0 are deemed more reliable. The QMEAN scores determined for *Cje*NANAS, *Cje*PseS and *Cje*LegS were -0.23, -0.90 and -0.66 respectively. Furthermore, as shown in Figures 4.17B, 4.18B and 4.19B, the normalised QMEAN scores for each of the *C. jejuni* models are well within the expected range for proteins of their respective molecular weights. The reliability of a generated model is strongly influenced by the sequence similarity to the template structure. The sequence identity of *Cje*NANAS, *Cje*PseS and *Cje*LegS with *Nme*NANAS is 58.6 %, 32.4 % and 34.1 % respectively. The best QMEAN score was observed for the *Cje*NANAS model which has the highest sequence identity to the template, and the worst QMEAN score was observed for the *Cje*PseS model which has the lowest sequence identity to the template.

Displayed in Figures 4.17C, 4.18C and 4.19C are plots of predicted C-α error against residue number for each of the three *C. jejuni* models. For each of the models, the C-α error is consistent between chains A and B, indicating that the quality of the model is not influenced by the quaternary state. Shown in Figure 4.17C, the C-α error for the *Cje*NANAS model is relatively consistent along the entire length of sequence when compared with the C-α error for the *Cje*PseS and *Cje*LegS models in Figures 4.18C and 4.19C respectively. The consistently low C-α error observed for the *Cje*NANAS model is likely due to the high level of similarity observed between the *Cje*NANAS and template sequences.

Conversely, in each of the local model quality plots for *Cje*PseS and *Cje*LegS there is a large spike in C alpha error observed around residues 55-75, and a second, smaller fluctuation around residues 270-290 (Figures 4.18C and 4.19C). These fluctuations correlate with the regions of low sequence identity observed between the SAS clades, as outlined in Section 4.2.2. Since the template used for modelling both *Cje*PseS and *Cje*LegS was a NANAS, it is unlikely that the aforementioned regions would be modelled accurately. The first of these regions (residues 55-75) corresponds to the β2-α2 loop, which was observed to have significant flexibility in the MD analysis for wild-type *Nme*NANAS (Chapter 2).

Displayed in Figures 4.20-4.22 are the model active sites of *Cje*NANAS, *Cje*PseS and *Cje*LegS respectively. Each model is overlaid with the active site residues from the *Nme*NANAS 1XUZ structure to highlight the differences and similarities between the model and template. Depicted in Figure 4.20, the active site residues of *Cje*NANAS are completely conserved with the template *Nme*NANAS structure. Given the degree of conservation between these two enzymes, it can be assumed that the interactions between the substrates and active site are analogous. This correlates with the results attained via kinetic analysis, where catalytic parameters were determined to be similar between *Cje*NANAS and *Nme*NANAS (Section 4.4.1).

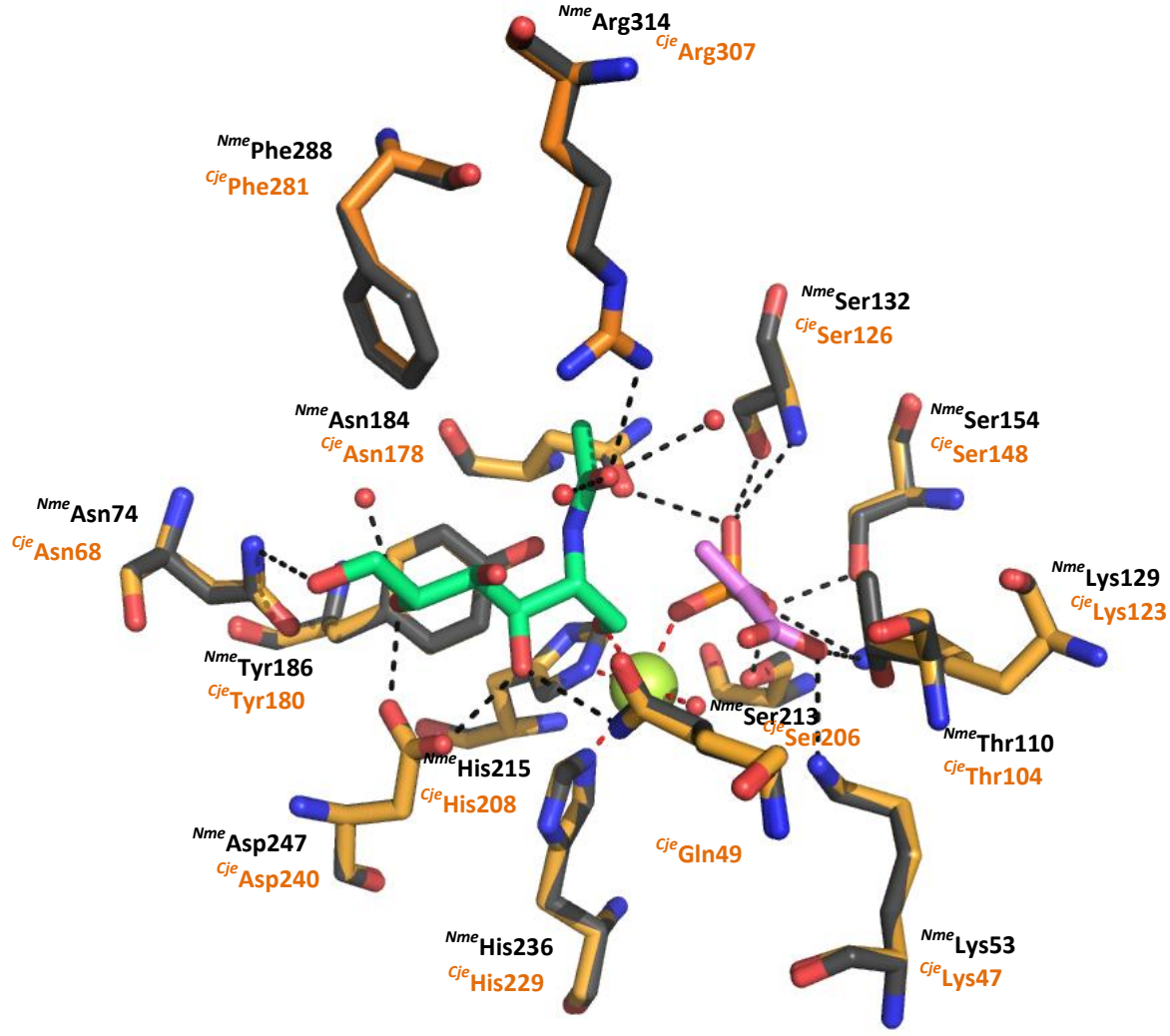


Figure 4.20 Overlay of *Nme*NANAS (crystal structure) and *Cje*NANAS (model) active site residues. Residues from *Nme*NANAS are coloured grey and *Cje*NANAS orange. Ligands are included from the *Nme*NANAS structure with Mn^{2+} in yellow, PEP in pink, rManNAc in green and waters as red spheres. Polar contacts are depicted with black dashes and metal-ligand interactions with red dashes. Crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*).

The majority of the active site residues in the *Cje*PseS model are conserved between *Nme*NANAS 1XUZ (Figure 4.21). Each of the metal binding residues are retained, as are each of the PEP binding residues with the exception of ^{Nme}Asn184. ^{Nme}Asn184 is only conserved in the NANAS clade (Section 4.2.2) and is replaced by ^{Cje}Ser181 in *Cje*PseS which is conserved within the PseS clade. The conserved sugar binding residues are similarly positioned however ^{Cje}Arg309 is slightly askew relative to ^{Nme}Arg314, possibly to accommodate residue ^{Cje}Arg282. Notably ^{Nme}Phe288 is replaced by ^{Cje}Arg282 in *Cje*PseS. ^{Cje}Arg282 is poised above the sugar substrate and possibly interacts with the second *N*-acetyl group of the PseS substrate. ^{Nme}Phe288 is fully conserved in the NANAS clade whilst ^{Cje}Arg282 is fully conserved in the PseS clade. ^{Nme}Asn74 is replaced by ^{Cje}Gly68 in *Cje*PseS, though the positioning of this residue may not be accurate. ^{Cje}Gly68 is contributed by the β2-α2 loop and, as shown in Figure 4.20B, is poorly modelled due to low sequence identity with the template.

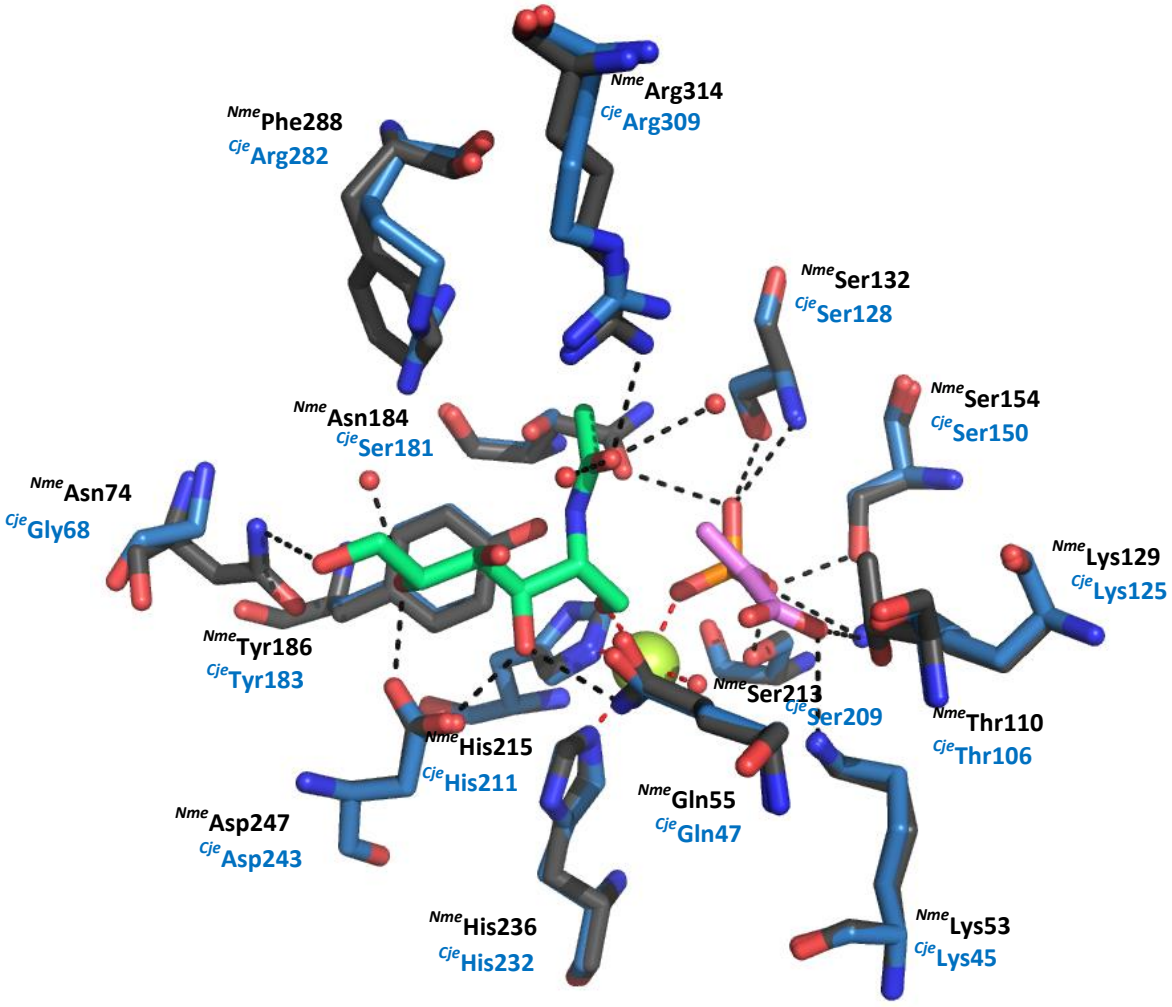


Figure 4.21 Overlay of *Nme*NANAS (crystal structure) and *Cje*PseS (model) active site residues. Residues from *Nme*NANAS are coloured grey and *Cje*PseS blue. Ligands are included from the *Nme*NANAS structure with Mn^{2+} in yellow, PEP in pink, rManNAc in green and waters as red spheres. Polar contacts are depicted with black dashes and metal-ligand interactions with red dashes. Crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*).

Displayed in Figure 4.24, most of the active site residues are similarly placed between the *Cje*LegS model and *Nme*NANAS 1XUZ. Similar to *Cje*PseS, each of the metal binding residues are retained, as are each of the PEP binding residues with the exception of ^{Nme}Asn184 which is replaced by ^{Cje}Thr175. ^{Cje}Thr175 is a highly conserved threonine in the LegS clade. All of the conserved sugar binding residues are similarly positioned in *Cje*LegS. ^{Nme}Phe288 is replaced by ^{Cje}Ile278 in *Cje*LegS, which is fully conserved in the LegS clade. None of the residues from the β2-α2 loop in *Cje*LegS were modelled within the active site, however, the positioning of this loop is unlikely to be very accurate given its low identity to the template sequence (Figure 4.21B).

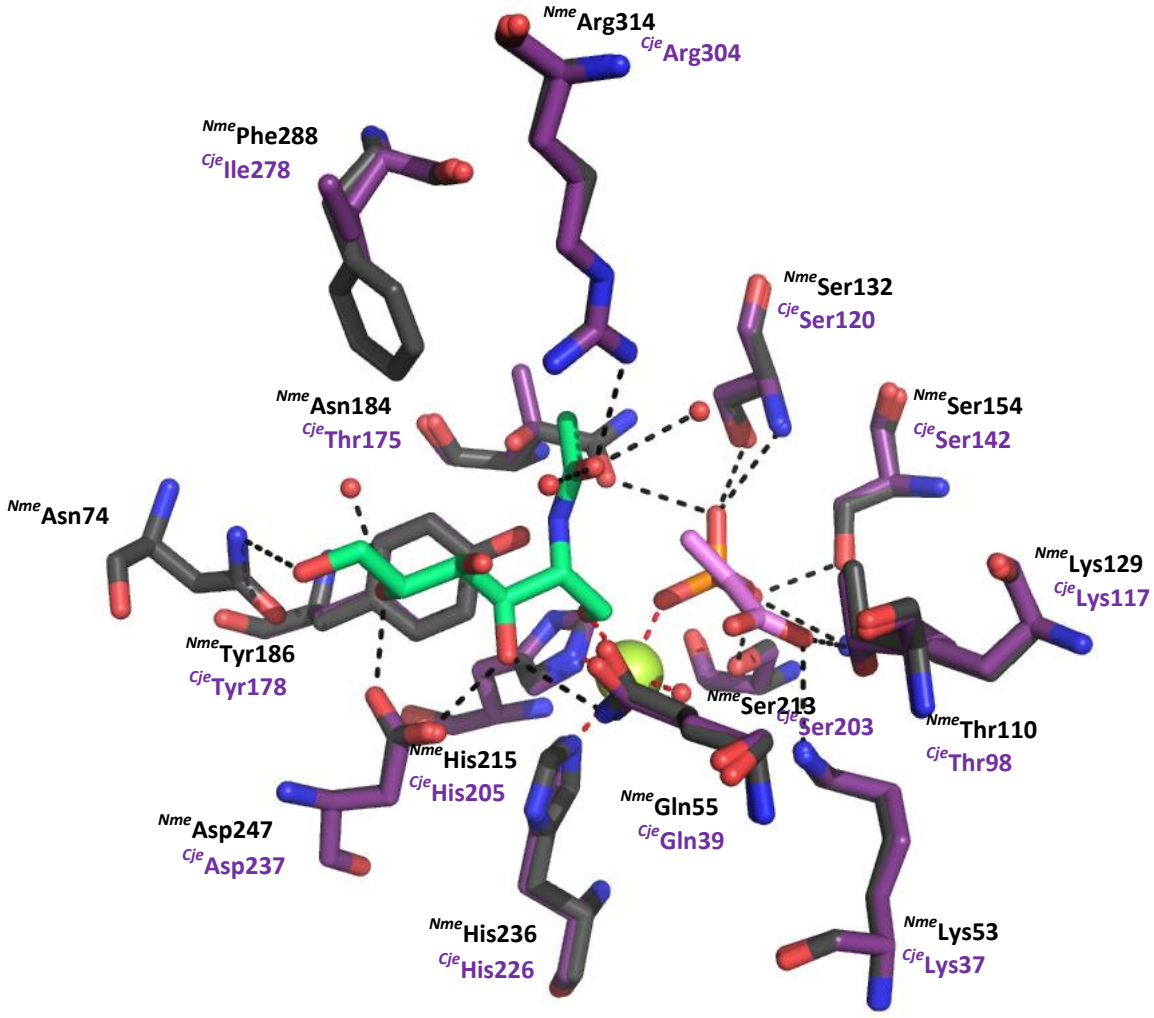


Figure 4.22 Overlay of *Nme*NANAS (crystal structure) and *Cje*LegS (model) active site residues. Residues from *Nme*NANAS are coloured grey and *Cje*LegS purple. Ligands are included from the *Nme*NANAS structure with Mn^{2+} in yellow, PEP in pink, rManNAc in green and waters as red spheres. Polar contacts are depicted with black dashes and metal-ligand interactions with red dashes. Crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*).

4.6.7 Substrate docking in *C. jejuni* SAS models

The active site residues of the *Cje*NANAS model are completely conserved with respect to the template *Nme*NANAS structure, and it can be subsequently assumed that the interactions between the sugar substrate and active site are almost identical (Section 4.6.6). Since the active site residues of the template are not fully conserved in either *Cje*PseS or *Cje*LegS and the sugar substrates differ, docking studies were used to infer the interactions within the respective active sites. Docking experiments were carried out by Dr Wanting Jiao for the *Cje*PseS model with Mn^{2+}, PEP and 2,4-diacetamido-2,4, 6-trideoxyaltrose as well as *Cje*LegS with Mn^{2+}, PEP and 2, 4-diacetamido-2, 4, 6-trideoxymannose.

Unfortunately, docking of 2,4-diacetamido-2,4,6-trideoxyaltrose in *Cje*PseS was unsuccessful. The PseS substrate did not dock in a suitable pose and several of the outcomes orientated the sugar molecule in a reversed conformation. Given that the enzyme has been confirmed to have PseS activity in the literature (*89*), the lack of a suitable reactive pose for the *Cje*PseS model with its substrate is likely an indicator that the model is critically flawed. The source of this flaw is probably the positioning of the *Cje*PseS β2-α2 loop which, as previously mentioned, has very low sequence identity to the *Nme*NANAS template. The lower sequence identity results in poor modelling of residues pertaining to that region, which in turn interferes with the docking of 2,4-diacetamido-2,4,6-trideoxyaltrose.

Conversely, 2,4-diacetamido-2,4,6-trideoxymannose was successfully docked into the active site of *Cje*LegS (Figure 4.23). The positioning of the metal and PEP binding residues are retained in the docking analysis of *Cje*LegS, as are their interactions. Both of the *N*-acetyl groups from the LegS substrate form interactions with ^{Cje}Arg304. ^{Cje}Arg304 is the congruent residue to ^{Nme}Arg314 from *Nme*NANAS. ^{Nme}Arg314 interacts with the *N*-acetyl functionality of ManNAc and is involved in steering the sugar into a reactive position with PEP (Chapter 2). Polar contacts are also formed between 2,4-diacetamido-2,4,6-trideoxymannose and ^{Cje}Glu39, ^{Cje}Tyr178 and ^{Cje}Asp237 which are each conserved ManNAc binding residues identified in *Nme*NANAS 1XUZ. Interestingly, no interactions are observed between 2,4-diacetamido-2,4,6-trideoxymannose and the β2-α2 loop of *Cje*LegS, however, this may be due to incorrect positioning of the β2-α2 loop in the *Cje*LegS model.

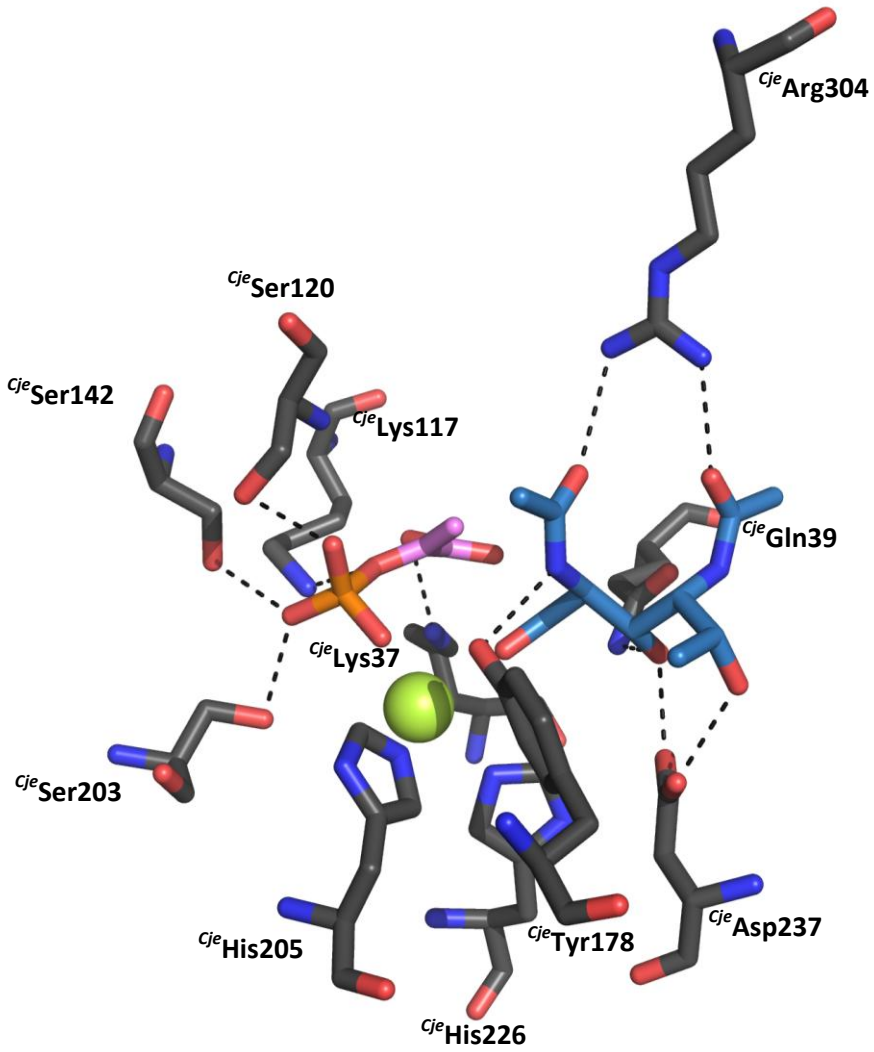


Figure 4.23 Best pose attained from docking of Mn^{2+}, PEP and 2, 4-diacetamido-2, 4, 6-trideoxymannose into the active site of the *Cje*LegS model. *Cje*LegS residues are shown in grey, with Mn^{2+} in yellow, PEP in pink and 2, 4-diacetamido-2, 4, 6-trideoxymannose in blue. Polar contacts are depicted with black dashes.

4.7 Cloning, expression and purification of *C. tepidum* SAS

4.7.1 Cloning of *C. tepidum* SAS

The gene encoding *Cte*SAS was codon optimised for expression in an *E. coli* cell line, synthesised and cloned into the vector pDONR221 by GeneArt®. DNA sequence encoding an N-terminal TEV protease cleavage site was incorporated during the construction of the *Cte*SAS gene. The *Cte*SAS gene was sub-cloned into the expression vector pDEST-17 which encodes an N-terminal His-tag.

4.7.2 Expression and purification

The vector pDEST-17 (*Cte*SAS) was transformed into three different expression cell lines to optimise the yield of soluble protein. Vector pDEST-17 (*Cte*SAS) was transformed into *E. coli* BL21 (DE3) Star cells, *E. coli* Rosetta™ 2 (DE3) cells and Chaperone 3 cells. In order to compare induced against non-induced samples, two cultures of each cell line were grown overnight in LB, with one induced by IPTG and the other not. The best yield of soluble protein was attained using Chaperone 3 cells as analysed by SDS PAGE. *Cte*SAS was thus expressed and purified from this cell line.

*Cte*SAS was expressed and purified using a similar method to that used for *Nme*NANAS wild-type as outlined in Section 8.4. It was determined that *Cte*SAS was unstable in 30 mM triethanolamine and in the absence of Mn^{2+} and PEP. The buffer used for SEC and storage of *Cte*SAS was thus modified to 20 mM bis-tris propane, 150 mM NaCl, 100 µM Mn^{2+} and 0.5 mM PEP at pH 7.5. Purified *Cte*SAS was concentrated, flash-frozen and stored at -80 °C. Samples of purified enzyme were analysed by SDS PAGE (Figure 4.24).

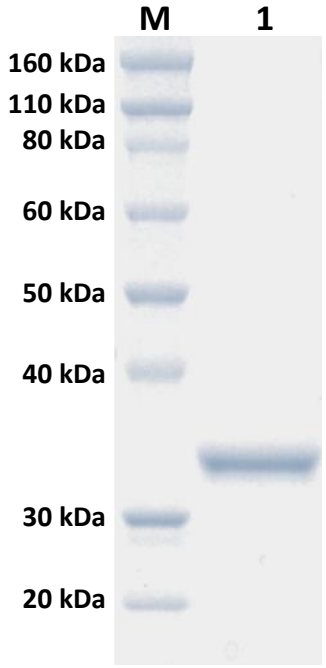


Figure 4.24 SDS-PAGE gel of purified *Cte*SAS. Lane M contains marker with molecular weights defined. Lane 1 contains purified *Cte*SAS.

4.8 Kinetic analysis of *C. tepidum* SAS

4.8.1 Attempted kinetic characterisation of *C. tepidum* SAS

To determine the kinetic parameters of *Cte*SAS, the consumption of PEP was monitored using the assay described in Section 8.5.1. *Cte*SAS was assayed with PEP and ManNAc at several different pH levels and temperatures as well as a variety of metal ions, however activity could not be observed. It has been previously proposed that *Cte*SAS is closely related to the mammalian SAS enzyme NANA-9-PS (*122*). It was hypothesised that *Cte*SAS could thereby utilise *N*-acetylmannosamine-6-phosphate (ManNAc-6-P) as its sugar substrate. ManNAc-6-P was thus tested for activity with PEP and ManNAc-6-P but no loss of PEP was observed. Lack of activity equates to a k_{cat} for the enzyme less than 0.004 s^{-1}.

As outlined in Section 4.2.2, *Cte*SAS belongs to a distinct clade of SASs (Type IV) from which no enzyme has been currently characterised. In an attempt to determine the sugar substrate for Type IV SASs, a range of different sugar substrates were assayed with *Cte*SAS. The ManNAc analogues outlined in Table 4.3 were each tested, however no detectable activity could be measured even in the presence of excess enzyme. Since neither substrate for Pse or Leg biosynthesis was available for testing, the ability of *Cte*SAS to catalyse either of these reactions cannot be currently ruled out.

4.9 Physical analysis of *C. tepidum* SASs

4.9.1 Circular dichroism

CD was used to assess and compare the secondary structure of *Cte*SAS with *Nme*NANAS (Figure 4.25). The K2D3 server (*142*) was used to estimate the secondary structure composition from CD spectra. Shown in Table 4.9, estimated secondary structure is compared to values derived from the crystal structure of *Nme*NANAS (PDB code: 1XUZ).

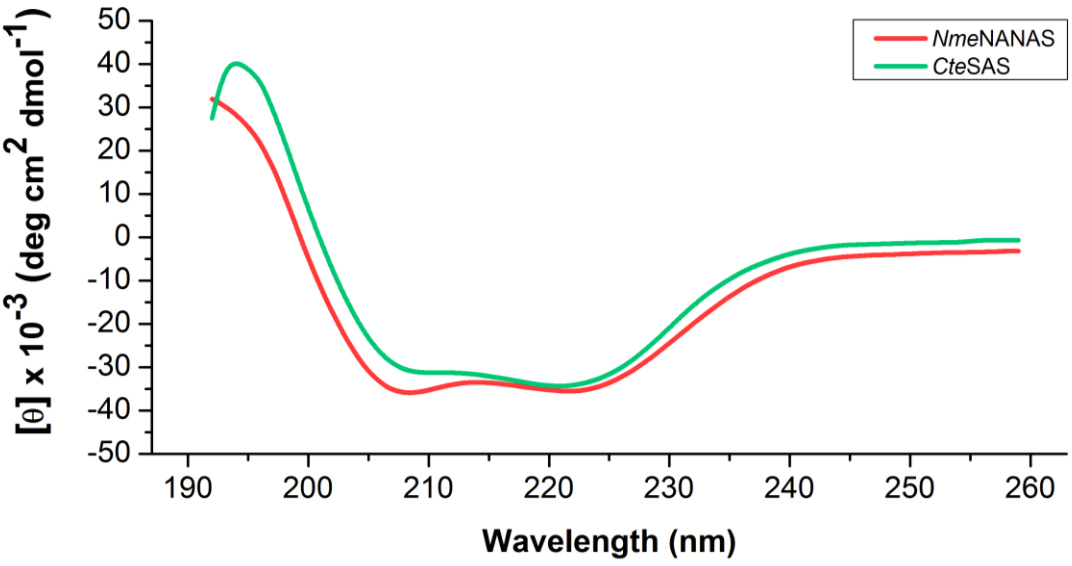


Figure 4.25 CD spectra for *Nme*NANAS (red) and *Cte*SAS (green).

	***Nme*NANAS**	***Nme*NANAS (structure)**	***CteSAS***
α-helix	59.9 %	50.3 %	57.5 %
β-sheet	12.3 %	15.5 %	18.6 %
other	27.8 %	34.2 %	23.9 %

Table 4.9 Predictions of secondary structure composition for *Nme*NANAS and *Cte*SAS calculated from CD data using K2D3 server (*142*). Secondary structure composition of *Nme*NANAS structure calculated from crystallographic coordinates 1XUZ.

As previously mentioned in Section 4.8, the activity of *Cte*SAS could not be validated. It was therefore neccessary to determine that *Cte*SAS was folded in solution using CD (Figure 4.25). As depicted in Table 4.9, *Cte*SAS has comparable levels of α-helix and β-sheet to *Nme*NANAS which is unexpected given that *Cte*SAS lacks an AFPL domain. However, as portrayed in Figure 4.25, the CD profile of *Cte*SAS is slightly different to that for *Nme*NANAS, which may be more indicative of structural differences between these two enzymes.

4.9.2 Mass spectrometry

The molecular mass of His_6-tagged *Cte*SAS was analysed using M. A theoretical mass of 34,112 Da was calculated for *Cte*SAS from the peptide sequence using ProtParam (*141*). The experimental mass determined by mass spectrometry was 34,111 Da which was almost identical to the aforementioned calculated mass.

4.9.3 Thermal stability

The effects of Mn^{2+}, PEP and ManNAc/rManNAc on the thermal stability of *Cte*SAS were assessed by DSF using the method outlined in Section 8.5.3. *Cte*SAS was treated with 10 mM EDTA prior to SEC to ensure removal of trace metals. The melting points determined for each sample were plotted graphically and compared with data obtained for *Nme*NANAS wild-type (Figure 4.26 and Table 4.10).

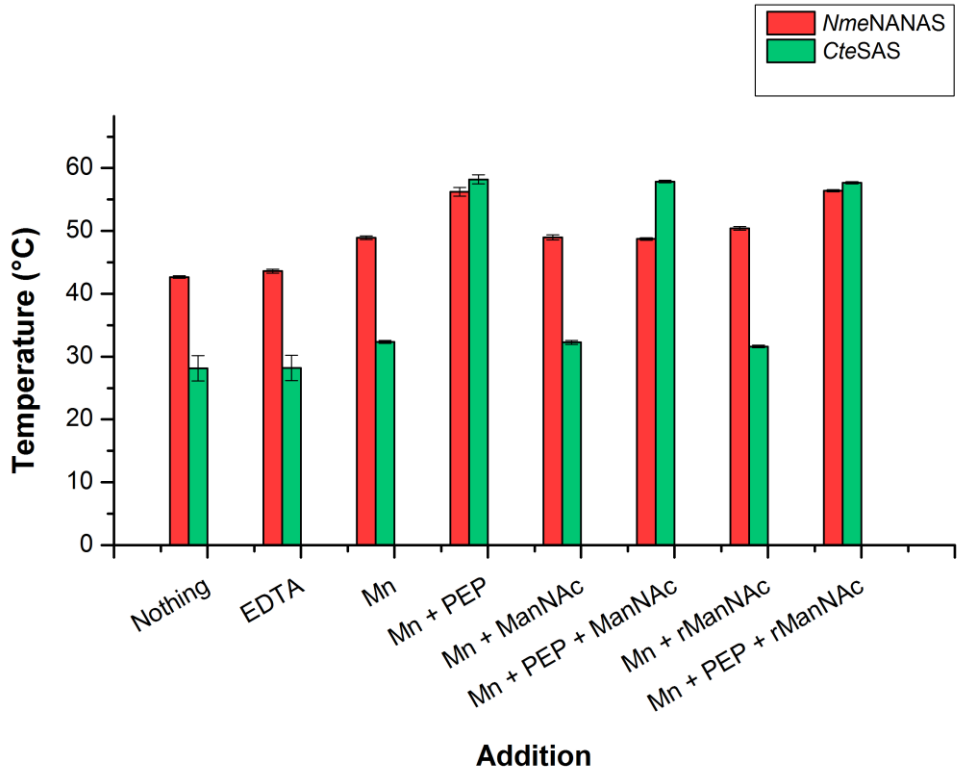


Figure 4.26 Bar graph of thermal melting points for *Nme*NANAS (red), and *Cte*SAS (green) in the presence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc at 10 mM. ManNAc/rManNAc was used at 1 mM when in combination with PEP.

Addition	*Nme*NANAS (°C)	*Cte*SAS (°C)
Nothing	42.7 ± 0.2	28.1 ± 1
EDTA	43.6 ± 0.3	28.2 ± 1
Mn^{2+}	48.9 ± 0.3	32.3 ± 0.2
Mn^{2+} + PEP	56.2 ± 0.7	58.2 ± 0.7
Mn^{2+} + ManNAc	49.0 ± 0.4	32.3 ± 0.3
Mn^{2+} + PEP + ManNAc	48.7 ± 0.2	57.8 ± 0.2
Mn^{2+} + rManNAc	50.4 ± 0.3	31.6 ± 0.2
Mn^{2+} + PEP + rManNAc	56.4 ± 0.2	57. 7 ± 0.2

Table 4.10 Thermal melting points for *Nme*NANAS and *Cte*SAS in the presence and absence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc at 10 mM. ManNAc/rManNAc was used at 1 mM when in combination with PEP.

Apoenzyme *Cte*SAS was determined to unfold at 28.1 ± 1 °C (Figure 4.26 and Table 4.10), and is far less stable than *apo*-*Nme*NANAS, which melts at 42.7 ± 0.2 °C. The addition of 1 mM EDTA to *apo*-*Cte*SAS did not influence the thermal stability of the enzyme which likely indicates the successful removal of trace metal ions during purification. The addition of 1 mM Mn^{2+} increased the melting point of *Cte*SAS, by 4 °C, similar to what is observed with *Nme*NANAS. An enormous thermal shift is observed for *Cte*SAS in the presence of both Mn^{2+} and PEP. The combined addition of 1 mM Mn^{2+} and 1 mM PEP to *Cte*SAS increased the melting point of apoenzyme by approximately 30 °C. The combined stabilisation of Mn^{2+} and PEP results in a melting point of 58.2 ± 0.7 °C for *Cte*SAS, and is slightly higher than that for *Nme*NANAS which melts at 56.2 ± 0.7 °C under the same conditions. It can be concluded that *Cte*SAS is extremely unstable in the absence of Mn^{2+} and PEP.

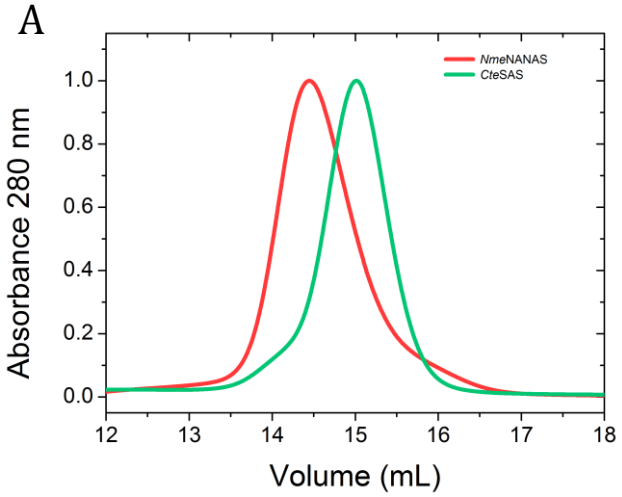
*Cte*SAS in the presence of Mn^{2+} and ManNAc/rManNAc has a similar melting temperature to *Cte*SAS in the presence of Mn^{2+}. This trend is observed in each of the SAS wild-type enzymes reported in this thesis. *Cte*SAS in the presence of Mn^{2+}, PEP and ManNAc/rManNAc has a similar melting temperature to *Cte*SAS in the presence of Mn^{2+} and PEP. As mentioned in Section 4.8, *Cte*SAS does not catalyse the condensation reaction between PEP and ManNAc. It is likely that the thermal shift observed for *CteSAS* in the presence of Mn^{2+}, PEP and ManNAc/rManNAc is caused by un-reacted PEP still present in the active site. This is analogous to the trend observed in *Cje*PseS (outlined in Section 4.4.2) which also does not catalyse NANA formation.

As previously mentioned, PEP binding in *Nme*NANAS pre-organises the binding site of ManNAc (Chapter 2). The mechanism by which this occurs involves a hydrogen bonding network which links the catalytic and AFPL domains. Interestingly, a significant thermal shift is observed in *Cte*SAS in the presence of PEP indicating a major structural change associated with PEP binding. The thermal shift upon PEP binding is observed in each of the wild-type SASs investigated in this thesis. However, *Cte*SAS does not have an AFPL domain nor does it contain the H-bond network residues conserved in each of the AFPL domain bearing SAS clades. It is possible that a congruent method for sugar binding site pre-organisation is employed by *Cte*SAS. Unfortunately due to the instability of *Cte*SAS without PEP over long periods, ITC experiments could not be used to accurately investigate PEP binding.

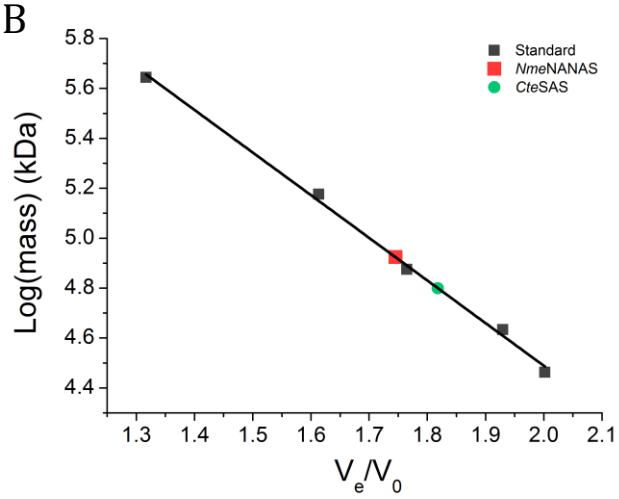
4.9.4 Analytical size exclusion

The quaternary state of *Cte*SAS was investigated via analytical SEC using the method outlined in Section 8.5.6. A calibration curve was generated using protein standards of known molecular weight. The elution volume (V_e) of protein standards relative to the column void volume (V_0) was plotted against the log of the protein standards molecular mass. A linear fit of the analytical SEC calibration data was determined to have an R^2 value of 0.987. The calibration curve was used to extrapolate the molecular masses of *Cte*SAS as shown in Figure 4.27.

A



B



C

	***Nme*NANAS**	***Cte*SAS**
V_e/V_0	1.74	1.82
Estimated mass (kDa)	81	61
Theoretical mass of dimer (kDa)	77	62

Figure 4.27 (A) Normalised analytical SEC chromatograms for *Nme*NANAS (red) and *Cte*SAS (green). (B) Analytical SEC calibration curve. Protein standards are marked in grey, *Nme*NANAS in red and *Cte*SAS in green. A line of best fit is plotted in black. (C) Table of estimated and theoretical masses for *Nme*NANAS and *Cte*SAS.

As shown in Figure 4.27 , *Cte*SAS elutes as a single species at a molecular weight of 61 kDa. The molecular weight of *Cte*SAS determined from analytical SEC is almost identical to the theoretical mass for dimeric *Cte*SAS. Each of the wild-type SASs examined in this thesis are also dimeric. Interestingly, *Cte*SAS elutes as a dimer even though it lacks an AFPL domain. As mentioned in Chapter 3, the AFPL domain was shown to be necessary in stabilising the dimer of *Nme*NANAS. The *Nme*NANAS AFPL domain truncation *Nme*G272Term was shown to have a concentration dependent monomer-dimer equilibrium through results from both analytical SEC and AUC.

4.9.5 Analytical ultracentrifugation

As discussed in Chapter 3, the AFPL domain plays an important role in stabilising the dimeric state of *Nme*NANAS. *Cte*SAS does not bear an AFPL domain and thus in order to elucidate its quaternary state, *Cte*SAS was examined through sedimentation velocity experiments. The method used for anlaysis is outlined in Section 8.5.7. To determine if the quaternary state of *Cte*SAS was concentration dependent, samples were run at 0.1 mg/mL, 0.5 mg/mL and 1.0 mg/mL. Displayed in Figure 4.28, data was fitted to a continuous size-distribution (c_s) model using the program SEDFIT (*150, 151*). Weight average molecular masses were calculated using weight average sedimentation coefficients and f/f_0 values from c_s fits.

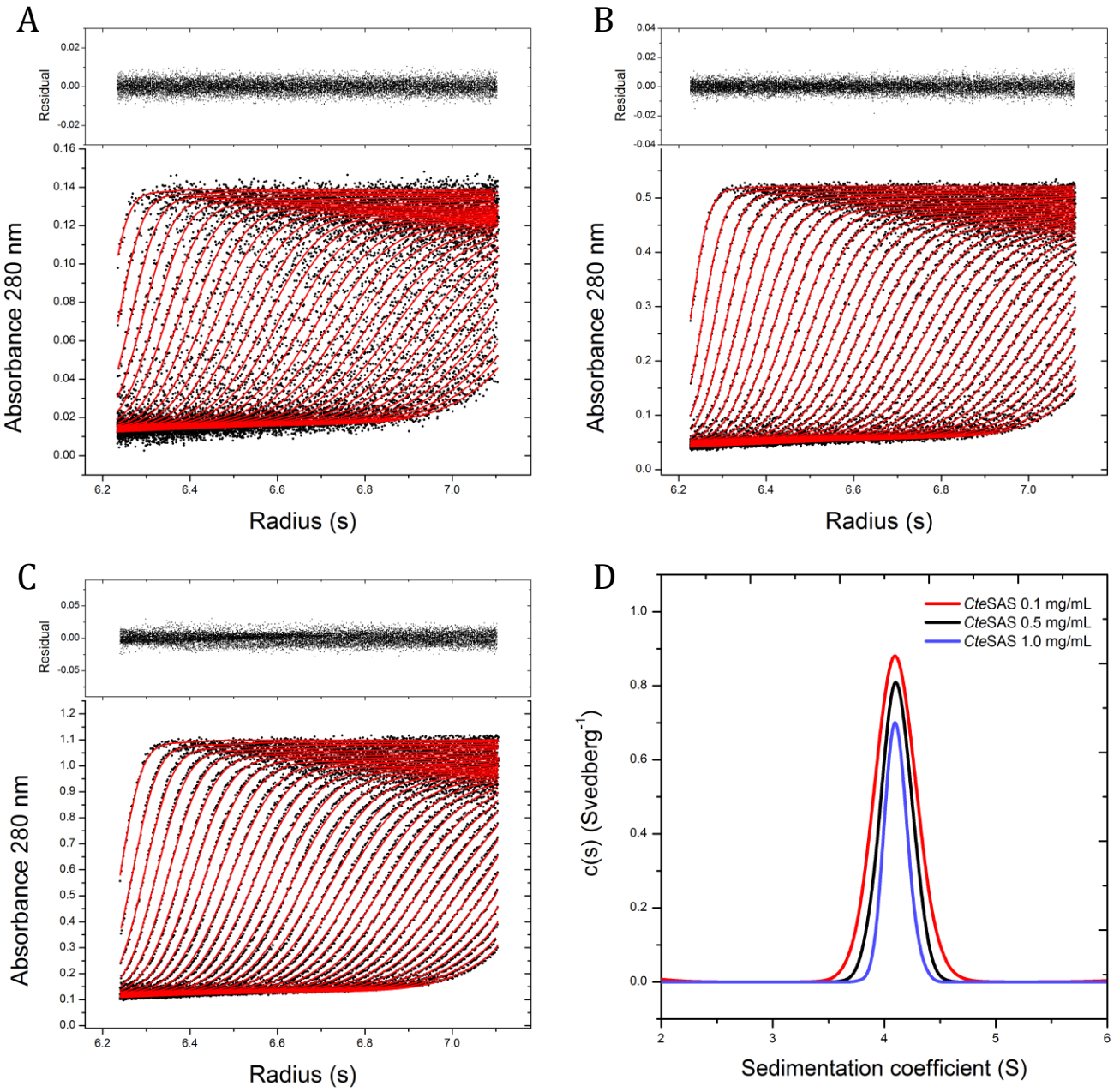


Figure 4.28 Analytical ultra centrifuge fits measured at 280 nm absorbance. (A) *Cte*SAS at 0.1 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (B) *Cte*SAS at 0.5 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (C) *Cte*SAS at 1.0 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (D) Normalised c(s) profiles for enzyme at 0.1 mg/mL (red), 0.5 mg/mL (black) and 1.0 mg/mL (purple).

Wild-type *Cte*SAS was determined to be a single species at each of the tested concentrations (Figure 4.28). The estimated molecular weight from AUC for *Cte*SAS wild-type was 57 kDa, which is closest to the molecular weight of dimeric *Cte*SAS. The results from AUC thereby correlate with those obtained from analytical SEC detailed in Section 4.9.5. As revealed by sedimentation velocity, *Cte*SAS forms a very strong dimer and does not show any inclination of a schizophrenic quaternary state even at 0.1 mg/mL.

Although the AFPL domain provides quaternary stabilisation in *Nme*NANAS, no such interaction appears necessary in maintaining the dimeric state of *Cte*SAS. Although the enzymes lack sequence for an AFPL domain, it should be noted that *Cte*SAS and other members of the Type IV SAS clade contain a short C-terminal extension (Figure 4.3). This extension may aid the quaternary stabilisation of *Cte*SAS, however, further analysis would be required to confirm this.

4.9.6 Attempted crystallisation

Several attempts at crystallising *Cte*SAS were made. *Cte*SAS was screened using conditions modified from those previously described for crystallising *Nme*NANAS (*112*). Protein crystallisation conditions were also screened using commercially available Morpheus, PACT and JCSG screens. Protein was concentrated to 6 mg/mL and 15 mg/mL and screened in the absence and presence of ligands. Screens were incubated at 5 °C and 20 °C for several months, but none of the conditions that were screened yielded crystals of diffraction quality. Modelling of *Cte*SAS was also attempted; however the sequence identity between *Cte*SAS and the *Nme*NANAS template was only 25.4 % which is well below the 30 % threshold for accurate modelling (*162*).

4.9.7 Small-angle X-ray scattering

SAXS was used to assess the solution structure of *Cte*SAS. *Cte*SAS was analysed in the absence and presence of 1 mM Mn^{2+} and 1 mM PEP (Figure 4.29). SAXS profiles of *Cte*SAS were determined using the methods outlined in Sections 8.5.8 and 8.5.9.

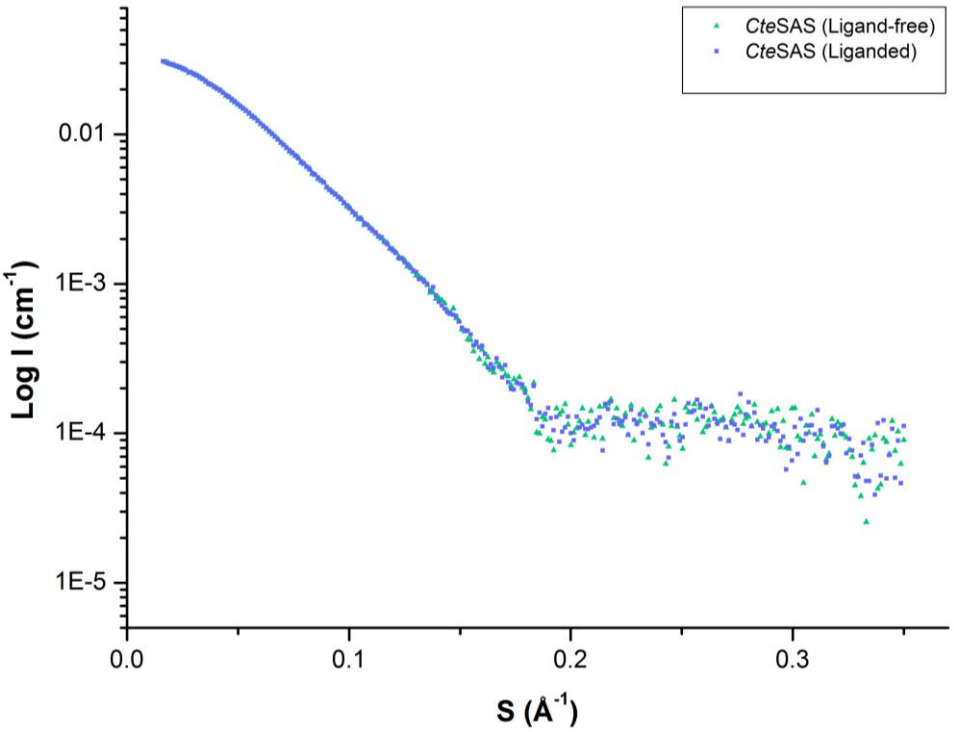


Figure 4.29 SAXS profiles of ligand-free *Cte*SAS (green) and *Cte*SAS in the presence of 1 mM Mn^{2+} and 1 mM PEP (blue). Scattering data was collected using approximately 10 mg mL^{-1} protein in 50 mM BTP at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

	***Cte*SAS (Ligand-free)**	***Cte*SAS (PEP bound)**
I_0 (cm^{-1})	0.033 ± 0.001	0.032 ± 0.001
R_g (Å)	30.1 ± 0.4	30.2 ± 0.3
D_{max} (Å)	103	105
Porod volume ($Å^3$)	110,161	110,671
Estimated mass (kDa)	68	69
Theoretical mass of dimer (kDa)	62	62

Table 4.11 SAXS structural parameters determined for *Cte*SAS. SAXS data obtained at a wavelength of 1.0332 Å.

The structure parameters obtained from SAXS are very similar between liganded and ligand-free *Cte*SAS (Table 4.11). The molecular masses of liganded and ligand-free *Cte*SAS were calculated from SAXS data using SAXS MoW (*154*), and were estimated to be 69 kDa and 68 kDa respectively. The experimental masses were similar to the theoretical molecular mass calculated for dimeric *Cte*SAS (Table 4.11). The results from SAXS indicate that *Cte*SAS is dimeric which acquiesces with data obtained from both analytical SEC and AUC.

Shown in Figure 4.29, the SAXS profile of *Cte*SAS is not altered in the presence or absence of ligand. As described in Chapter 3, wild-type *Nme*NANAS undergoes a conformational change in the presence of Mn^{2+} and PEP that is detectable by SAXS. The AFPL domain of *Nme*NANAS is involved in a hydrogen bonding network where PEP binding pre-organises the binding site of ManNAc. Due to the involvement of the AFPL domain in this process, a conformational change is not observed for *Nme*G272Term in the presence of ligand. It is likely that *Cte*SAS does not undergo a conformational change upon PEP binding due to the absence of an AFPL domain. Interestingly, a significant thermal stabilisation of *Cte*SAS is observed in the presence of PEP, as detailed in Section 4.9.3. Unfortunately, if there is a physical change in the structure of *Cte*SAS induced by PEP binding it cannot be observed via SAXS.

As mentioned in Section 4.9.6, neither a crystal structure nor model of *Cte*SAS could be obtained. The experimental scattering of *Cte*SAS was thus fitted to theoretical scattering generated from an AFPL domain truncation of the *Nme*NANAS crystal structure 1XUZ (*112*). The truncated *Nme*NANAS structure was also used to generate a CRYSOL fit for *Nme*G272Term as outlined in Chapter 3. Shown in Figure 4.30A, the experimental SAXS profile of liganded *Cte*SAS fits well with the CRYSOL generated profile for the *Nme*G272Term model and has a Chi value of 0.68. In comparison, the experimental SAXS profile of liganded *Nme*G272Term fits slightly better, with a Chi value of 0.52 (Figure 4.30B). The results from CRYSOL indicate that *Cte*SAS adopts a dimer of $(\beta/\alpha)_8$ barrels which correlates with the results from analytical SEC, AUC and sequence homology of *Cte*SAS to other members of the SAS family.

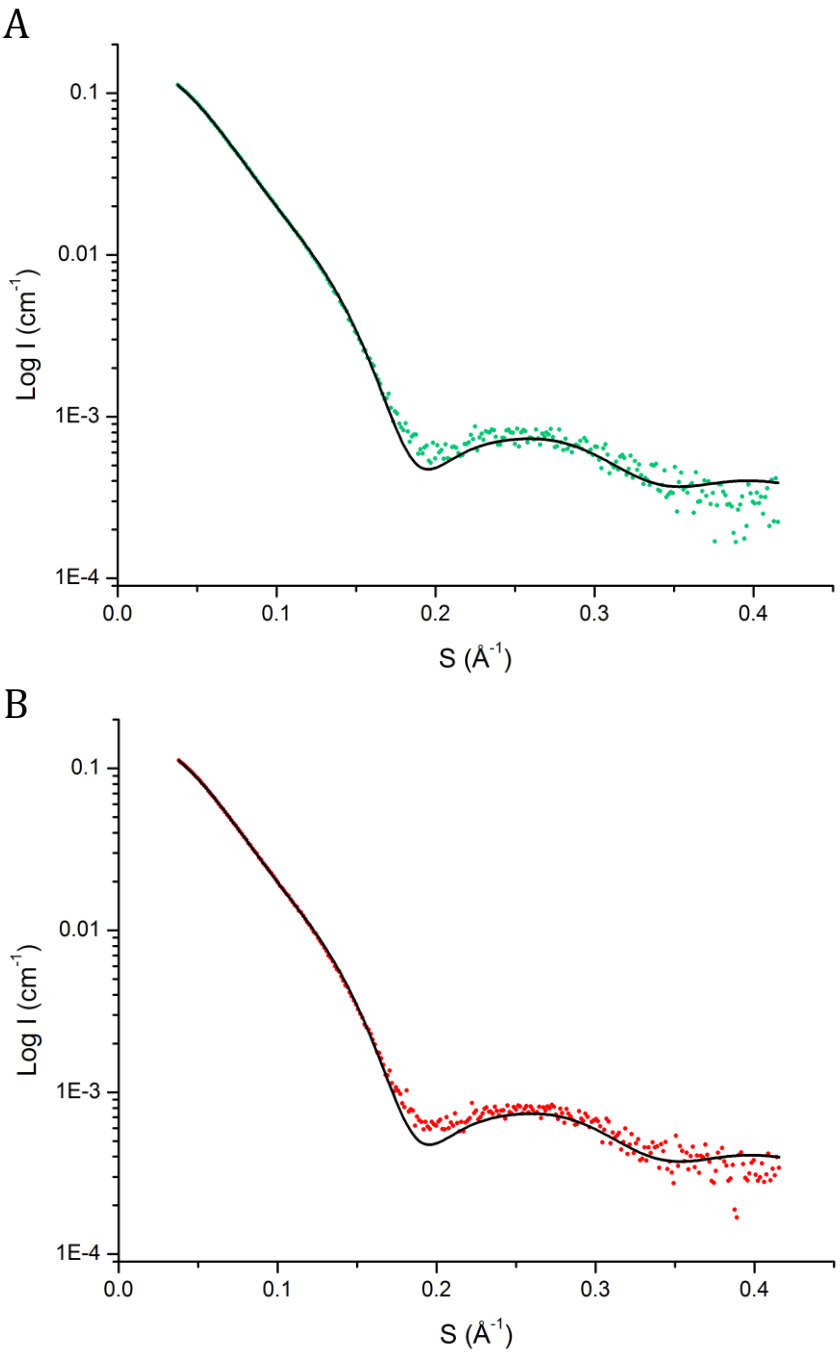


Figure 4.30 Comparisons of experimental and theoretical SAXS profiles using CRYSOL. (A) Comparison of SAXS profiles for *Cte*SAS in the presence of Mn^{2+} and PEP (green) with theoretical scattering generated from dimer model of *Nme*G272Term (black line). (B) Comparison of SAXS profiles for *Nme*G272Term in the presence of Mn^{2+} and PEP (red) with theoretical scattering generated from dimer model of *Nme*G272Term (black line). Theoretical scattering for *Nme*G272Term was generated from crystallographic coordinates 1XUZ. *Cte*SAS SAXS data was collected using approximately 10 mg mL^{-1} protein in 50 mM BTP at pH 7.5. *Nme*G272Term SAXS data was collected using approximately 10 mg mL^{-1} protein in 30 mM triethanolamine-HCl buffer at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

4.10 Summary of results

The three SAS paralogues from *C. jejuni* were successfully amplified from genomic DNA and cloned into His-tag expression vectors. *Cje*NANAS, *Cje*PseS and *Cje*LegS were each expressed and purified by metal affinity and size exclusion chromatography. NANAS activity was confirmed for *Cje*NANAS and comparable kinetic parameters to wild-type *Nme*NANAS were obtained. Low levels of NANAS activity were also observed with *Cje*LegS, however, no activity was yielded with *Cje*PseS. Activity with ManNAc analogues was not observed with any of the three *C. jejuni* SASs.

The thermal stabilities of *Cje*NANAS, *Cje*PseS and *Cje*LegS were analysed in the presence of various ligand combinations. Similar to the observations made with wild-type *Nme*NANAS, a significant thermal shift was noted upon the addition of PEP for each of the three *C. jejuni* SASs. The ITC profiles of *Cje*NANAS, *Cje*PseS and *Cje*LegS with Mn^{2+} and PEP were also analogous to *Nme*NANAS wild-type. As outlined in Chapter 3, PEP liganding pre-organises the ManNAc binding site via a hydrogen bonding network. Given the similarity in experimental observations for PEP binding it is highly likely that the binding mechanism of PEP in *Nme*NANAS is conserved in each of the *C. jejuni* SASs.

The molecular masses of *Cje*NANAS, *Cje*PseS and *Cje*LegS were successfully ascertained using mass spectroscopy and each of the enzymes were determined to be folded in solution using CD. Each of the three *C. jejuni* paralogues was confirmed to be dimeric by SAXS and analytical SEC. Although extensive attempts to crystallise the enzymes were made, no diffraction quality crystals for *Cje*NANAS, *Cje*PseS or *Cje*LegS were obtained.

In an attempt to glean the active site interactions of the *C. jejuni* SASs, models of each of the enzymes were generated using the SWISS-MODEL server. The structures were determined to have good overall homology to the template structure from *Nme*NANAS, with only small regions of the model lacking good sequence identity. Docking of the PseS substrate into the active site of the *Cje*PseS was unsuccessful, likely due to incorrect positioning of residues from the β2-α2 loop. In comparison, 2,4-diacetamido-2,4,6-trideoxymannose was successfully docked into the active site of *Cje*LegS, revealing several key interactions with active site residues.

The putative SAS from *C. tepidum* was successfully expressed and purified by metal affinity and size exclusion chromatography. Unfortunately no catalytic activity for *Cte*SAS was afforded with either ManNAc or any of the ManNAc analogues available. *Cte*SAS was confirmed to be folded in solution using CD and the molecular mass was obtained using mass spectroscopy.

The thermal stability of *Cte*SAS was analysed using DSF. Similar to each of the wild-type enzymes analysed in this thesis, PEP was observed to have a significant stabilising effect upon *Cte*SAS. Furthermore, *Cte*SAS was noted to be relatively unstable in the absence of PEP when compared with the other wild-type enzymes examined.

*Cte*SAS was proven to be dimeric though SAXS, analytical SEC and AUC, even though the enzyme lacks an AFPL domain. As discussed in Chapter 3, the AFPL domain plays an important role in stabilising the quaternary structure of wild-type *Nme*NANAS. Several attempts to crystallise *Cte*SAS were made, however, no diffraction quality crystals were obtained. Modelling of the *Cte*SAS structure was not possible due to low sequence identity to currently available template structures.

4.11 Discussion

The aim of the work outlined in Chapter 4 was to uncover the evolutionary relationships within the bacterial SAS family. The first step in the analysis was a sequence alignment of known and potential bacterial SAS sequences. A dendogram generated from the sequence alignment identified four distinct clades within the SAS family. These were denoted NANAS, LegS, PseS and Type IV SASs, the latter of which had unknown catalytic activity. The sequence alignment of bacterial SASs also identified regions of high sequence identity that were specific to each clade; including a region corresponding to the β2-α2 loop of *Nme*NANAS. Given that the majority of active site residues are otherwise conserved, the residues from this loop are likely involved in distinguishing between the different sugar substrates of NANAS, LegS, PseS and Type IV SASs.

Further investigation of the evolutionary relationships was undertaken by cloning, expressing and purifying a selected member from each clade. The NANAS, LegS and PseS from *C. jejuni* were selected for analysis as was the putative Type IV SAS from *C. tepidum*. NANAS activity was observed in *Cje*NANAS with comparable catalytic efficiency to wild-type *Nme*NANAS. Low levels of NANAS activity was also observed in *Cje*LegS, however, no NANAS activity was detected in either *Cje*PseS or *Cte*SAS. Several ManNAc analogues were assayed with each of the four SAS homologues, however, no activity was detected with any of the compounds tested.

DSF revealed that *Cje*NANAS, *Cje*PseS, *Cje*LegS and *Cte*SAS were each thermally stabilised in the presence of PEP, with ITC and SAXS analysis confirming that each *C. jejuni* enzyme had a similar profile to *Nme*NANAS when in the presence of PEP. The mechanism by which PEP pre-organises the sugar substrate binding site in *Nme*NANAS is potentially conserved in *Cje*NANAS, *Cje*PseS and *Cje*LegS. Interestingly, *Cte*SAS was stabilised by PEP even though it naturally lacks an AFPL domain. Conversely, the *Nme*NANAS truncation - *Nme*G272Term, was not stabilised by PEP which may indicate that *Cte*SAS employs a different rearrangement mechanism upon PEP binding when compared with AFPL domain bearing SASs. Each of the three *C. jejuni* SASs as well as *Cte*SAS were confirmed to be dimeric. Models generated of each of the *C. jejuni* SASs identified key structural differences and similarities, and successful docking of the LegS substrate into the active site of *Cje*LegS revealed several potential active site interactions that may be involved in substrate binding.

Chapter 5

Characterisation of *Streptococcus suis* sialic acid synthase

5.1 Background

5.1.1 *S. agalactiae* NANAS

The SAS from *Streptococcus agalactiae* has been previously characterised as an NANAS (*105*). *S. agalactiae* NANAS (*Sag*NANAS) is the only bacterial SAS which has shown the propensity to utilise alternative sugar substrates. *Sag*NANAS was determined to be most active with ManNAc, however trace levels of activity were observed using both GlcNAc and GalNAc as alternative substrates along with PEP (Table 5.1). *Sag*NANAS catalytic activity was characterised in these studies using a thiobarbituric acid stopped assay (*105*).

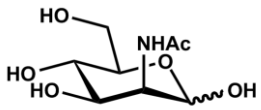
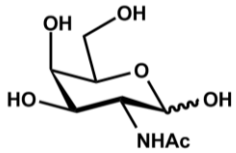
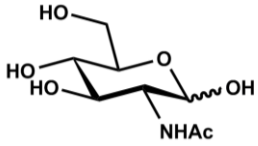
Name	Structure	Relative Activity (%)
N-acetyl-D-mannosamine		100
N-acetyl-D-galactosamine		15
N-acetyl-D-glucosamine		5

Table 5.1 Alternative sugar substrates utilised by *S. agalactiae* NANAS. Percentage relative activity of substrates is calculated with respect to activity with ManNAc (*105*).

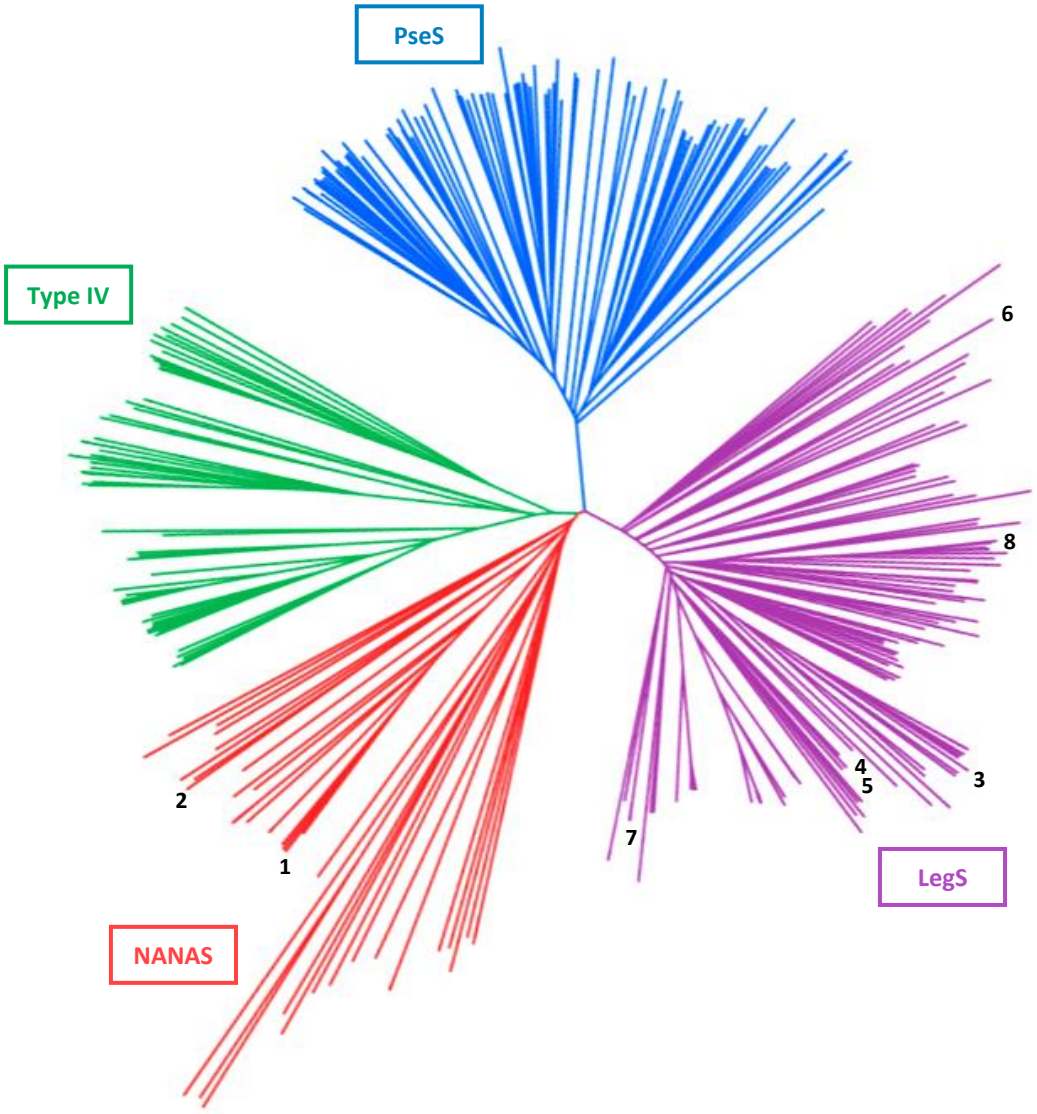
The melting point of *Sag*NANAS was determined to be around 37 °C (*105*). Maximal activity was observed at a pH of 7.0, however activity was observed between pH 5.0 and 10.0. The metal dependency of *Sag*NANAS was also analysed. Complete inhibition of the *Sag*NANAS enzyme was observed after treatment with EDTA, indicating an essential role of metal ions in catalysis. Maximal activity was observed in the presence of Co^{2+}, with Mn^{2+} being the second strongest activating metal ion. Both Cu^{2+} and Zn^{2+} were observed to have dramatic inhibitory affects upon *Sag*NANAS. *Sag*NANAS was also inhibited by phenylglyoxal suggesting that one or more arginine residues were essential to the catalytic function of the enzyme (*105*).

5.1.2 *Streptococcus suis*

Using the online BLAST server (*164*), the SAS from *Streptococcus suis* was identified as the enzyme with the greatest sequence similarity to *Sag*NANAS, sharing 81% identity. *S. suis* is a Gram-positive anaerobe and is primarily responsible for bacterial meningitis in pigs (*165*). Although the exact pathogenetic mechanism of *S. suis* is currently unknown, the bacterium has been linked to several syndromes in swine including arthritis, pneumonia, septicaemia, endocarditis, polyserositis, abortions and abscesses (*166, 167*). *S. suis* infection was estimated to be a 300 million dollar economic loss to the swine industry in the United States alone (*167*). *S. suis* infection is zoonotic (*165*), and can cause meningitis in humans resulting in permanent hearing loss, septicaemia, endocarditis and death (*168-171*). Deaths caused by *S. suis* infection occur primarily in abattoir workers and butchers handling infected pork (*168, 169*). Immunoprophylaxis of *S. suis* infection is currently unavailable and thus control of the disease is dependent upon antimicrobials such as penicillin (*172*).

5.1.3 Phylogeny of *S. suis* SAS

Described in Section 4.2.2, an extensive sequence alignment of 410 known and potential SAS genes was constructed using ClustalW2 (*155, 156*). Depicted in Figure 5.1 is a phylogram derived from the SAS sequence alignment. The SAS phylogenetic tree identifies four major clades within the SAS family. These clades have been assigned in this thesis as NANAS, LegS, PseS and a fourth grouping of enzymes with unknown activity labelled as Type IV.



	Organism	SAS activity
1	*Neisseria meningitidis*	NANAS (*5, 103, 112*)
2	*Campylobacter jejuni* (Cj1141)	NANAS (*90*)
3	*Escherichia coli*	NANAS (*95, 158*)
4	*Streptococcus agalactiae*	NANAS (*105*)
5	*Streptococcus suis*	NANAS
6	*Legionella pneumophila*	LegS (*68*)
7	*Campylobacter jejuni* (Cj1327)	LegS (*135*)
8	*Photobacterium profundum* strain 3TCK	LegS (*134*)

Figure 5.1 Phylogenetic tree of bacterial SASs. In red are NANAS, purple are LegSs, blue are PseSs and Type IV SASs. Numbers depict SASs which have been identified in the literature. Organism name and known activity is tabulated above.

As aforementioned, the SAS from *S. agalactiae* has been characterised as a NANAS (*105*). However, as shown In the SAS phylogram, it is grouped in the LegS clade (Figure 5.1). *Sag*NANAS and *Ssu*SAS share 51 % and 48 % sequence identity with *Cje*LegS respectively. In comparison, *Sag*NANAS and *Ssu*SAS share only 35 % and 34 % identity with *Nme*NANAS respectively. Interestingly, the SAS from *E. coli* has also been previously characterised as a NANAS (*95, 158*) yet is located in the LegS clade. Like *Sag*NANAS, *Eco*NANAS shares higher identity with *Cje*LegS (54 %) than with *Nme*NANAS (34 %).

As mentioned in Chapter 4, *Cje*LegS is capable of NANAS activity although it is significantly diminished when compared with *Cje*NANAS. It is possible that LegSs are capable of NANAS activity resulting in the mislabelling of the SASs from *S. agalactiae* and *E. coli* as NANASs. Furthermore, the sugar substrate for Leg synthesis (2,4-diacetamido-2,4,6-trideoxymannose) is both structurally and stereochemically similar to ManNAc (Figure 5.2). Unfortunately the activity of these enzymes with the substrate for Leg biosynthesis has not currently been tested.

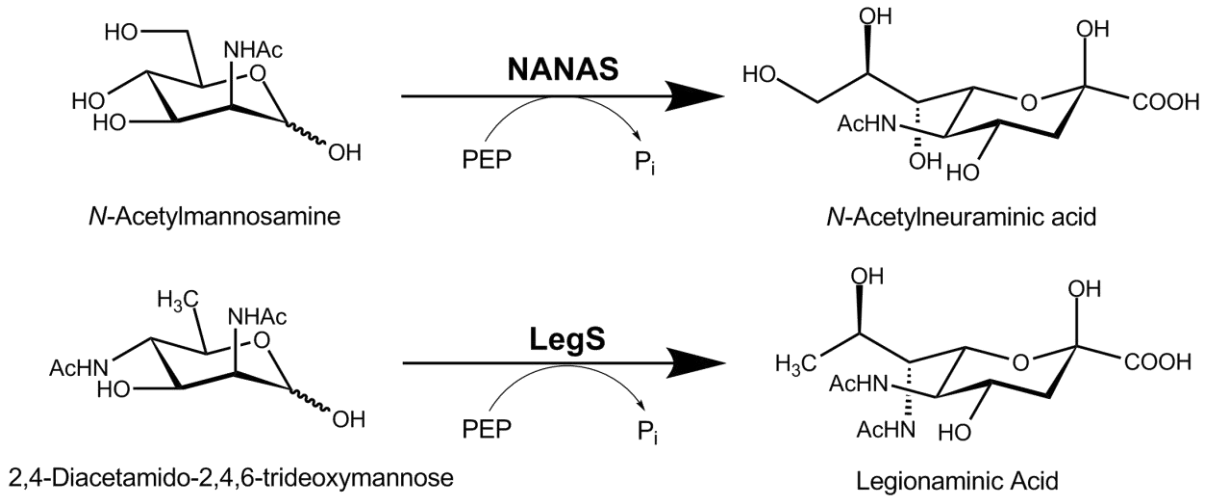


Figure 5.2 Enzymatic reactions catalysed by NANAS and LegS.

5.1.4 Sequence analysis of *Ssu*SAS

Displayed in Figure 5.3 is a sequence alignment of *Ssu*SAS, *Sag*NANAS, *Eco*NANAS and *Nme*NANAS with identified LegSs from *C. jejuni* and *L. pneumophila*. Figure 5.3 clearly illustrates the similarity of *Ssu*SAS, *Sag*NANAS and *Eco*NANAS to *Cje*LegS and *Lpn*LegS. It can also be observed that *Ssu*SAS, *Sag*NANAS, *Eco*NANAS, *Cje*LegS and *Lpn*LegS have regions which are significantly different to the *Nme*NANAS sequence, specifically the sequence pertaining to the β2-α2 loop which was implied in Chapter 4 to play a role in substrate recognition.

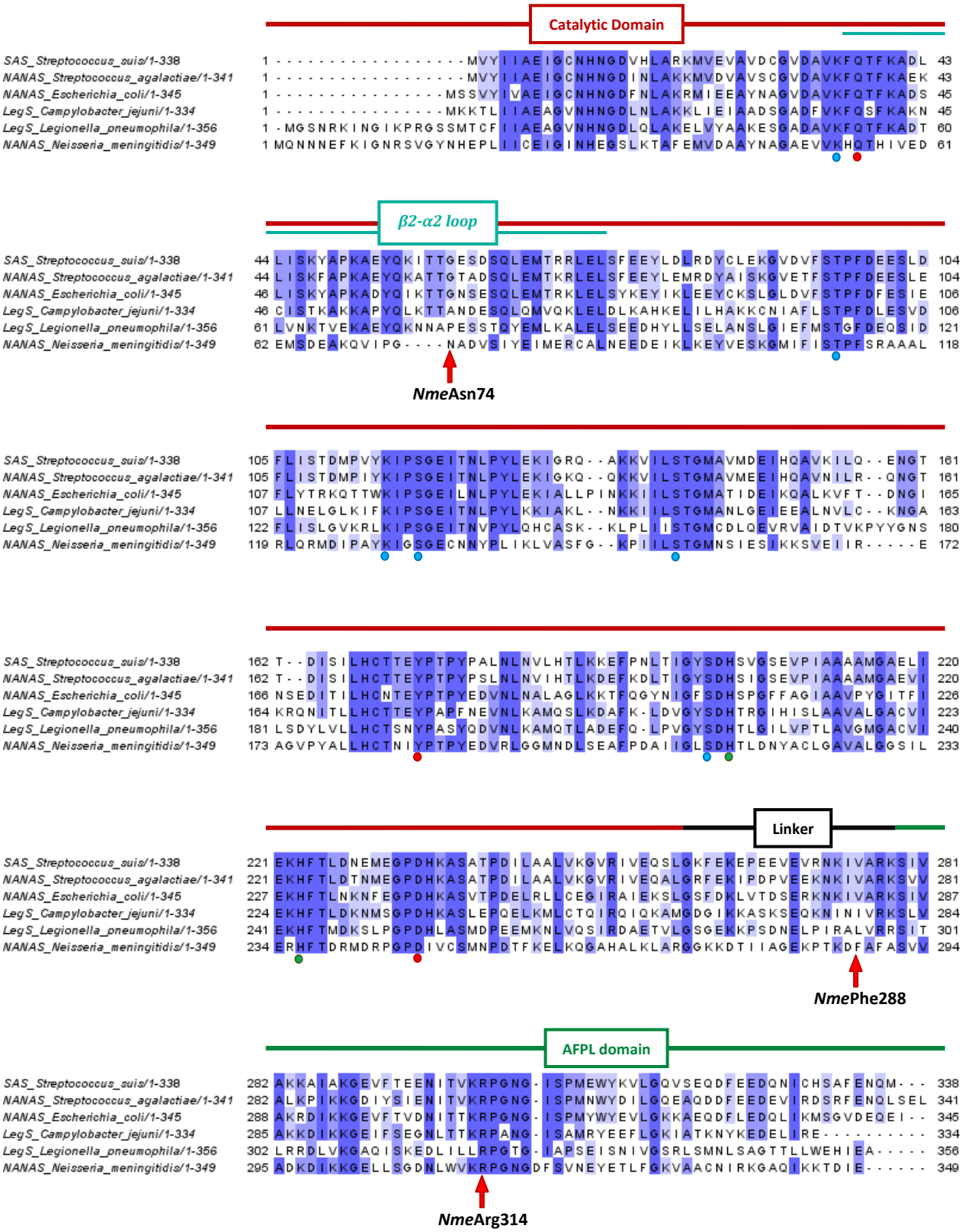


Figure 5.3 Sequence alignment of *Ssu*SAS, *Sag*NANAS, *Eco*NANAS, *Cje*LegS, *Lpn*LegS and *Nme*NANAS displayed using Jalview (*159*). Residues highlighted in blue indicate conservation. Conserved active site residues from *Nme*NANAS are marked with dots; metal binding residues in green, PEP binding residues in blue and ManNAc binding residues in red. Lines indicate sequence pertaining to *Nme*NANAS domains; catalytic barrel in red, β2-α2 loop in blue, linker in black and AFPL domain in green. The residue and domain annotation is based on the *Nme*NANAS crystal structure (PDB code (1XUZ) (*112*)).

All of the key active site residues identified in the *Nme*NANAS crystal structure are conserved in each of the aligned sequences with the exception of *Nme*Asn74 and *Nme*Phe288 (Figures 5.3 and 5.4). As explained in Chapter 4, *Nme*Asn74 and *Nme*Phe288 are only conserved in the NANAS clade. Interestingly, the sequences in the regions corresponding to *Nme*Asn74 and *Nme*Phe288 are similar between each of the enzymes aligned with the exception of *Nme*NANAS, further indicating that each of these enzymes are both similar to one another and distinct from *Nme*NANAS.

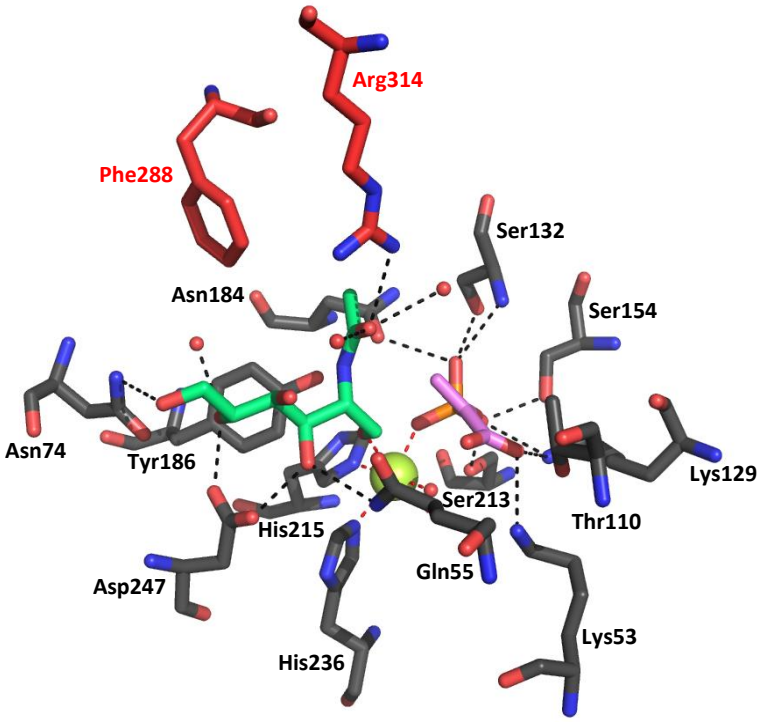


Figure 5.4 Close up of *Nme*NANAS active site. Residues contributed by the catalytic domain are shown in grey and residues from linker and AFPL domain in red. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are shown in red. H-bonds are depicted with black dashes and dipolar bonds with red dashes (PDB code (1XUZ) (*112*)).

5.1.5 Goals of this research

As previously mentioned, *Sag*NANAS is the only known bacterial SAS with the ability to utilise alternative sugar substrates (*105*). Activity with alternative sugar substrates was not observed in any of the SASs from *C. jejuni* nor the Type IV SAS from *C. tepidum* (Chapter 4). The SAS from *S. suis* was identified to have the greatest sequence similarity to *Sag*NANAS. *Ssu*SAS was thus expressed, purified and characterised in an effort to determine whether substrate promiscuity was unique to *Sag*NANAS or could be observed in other SASs.

5.2 Cloning, expression and purification of *Ssu*SAS

5.2.1 Cloning of *Ssu*SAS

The gene encoding *Ssu*SAS was codon optimised for expression in an *E. coli* cell line, synthesised and cloned into the vector pDONR 221 by GeneArt®. DNA sequence encoding an N-terminal TEV protease cleavage site was incorporated during the construction of the *Ssu*SAS gene. The *Ssu*SAS gene was sub-cloned into the expression vector pDEST-17 which encodes an N-terminal His-tag.

5.2.2 Expression and purification of *Ssu*SAS

The vector pDEST-17 (*Ssu*SAS) was transformed into three different expression cell lines to optimise the yield of soluble protein. Vector pDEST-17 (*Ssu*SAS) was transformed into *E. coli* BL21 (DE3) Star cells, *E. coli* Rosetta™ 2 (DE3) cells and Chaperone 3 cells. In order to compare induced against non-induced samples, two cultures of each cell line were grown overnight in LB, with one induced by IPTG and the other not. The best yield of soluble protein was attained using Chaperone 3 cells as analysed by SDS PAGE. *Ssu*SAS was thus expressed and purified from this cell line. *Ssu*SAS was expressed and purified using a similar method to that used for *Nme*NANAS wild-type as outlined in Section 8.4. Purified *Ssu*SAS was concentrated, flash-frozen and stored at -80 °C. Samples of purified enzyme were analysed by SDS PAGE (Figure 5.5).

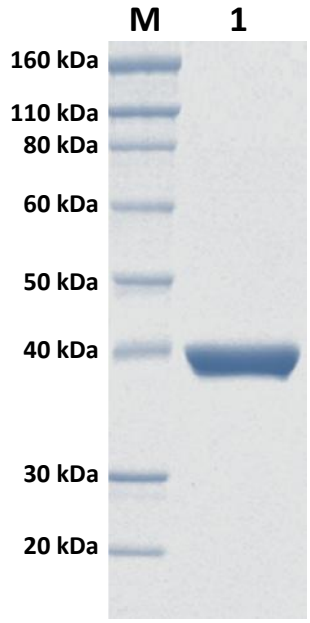


Figure 5.5 SDS-PAGE gel of purified *S. suis* SAS enzymes. Lane M contains marker with molecular weights defined. Lane 1 contains purified *Ssu*SAS.

5.3 Kinetic analysis of *Ssu*SAS

5.3.1 Michaelis-Menten kinetics of *Ssu*SAS

To determine the kinetic parameters of *Ssu*SAS, the consumption of PEP was monitored using the assay described in Section 8.5.1. Cuvettes containing 50 mM BTP (pH 7.5), 1 mM $MnCl_2$, PEP (varied) and ManNAc (varied) were incubated for 10 min at 25 °C. The enzymatic reaction was initiated by the addition of purified enzyme. Michaelis-Menten steady-state approximations were utilized to determine kinetic parameters (Figure 5.6). Kinetic parameters were compared to the values ascertained for *Nme*NANAS wild-type (Table 5.2).

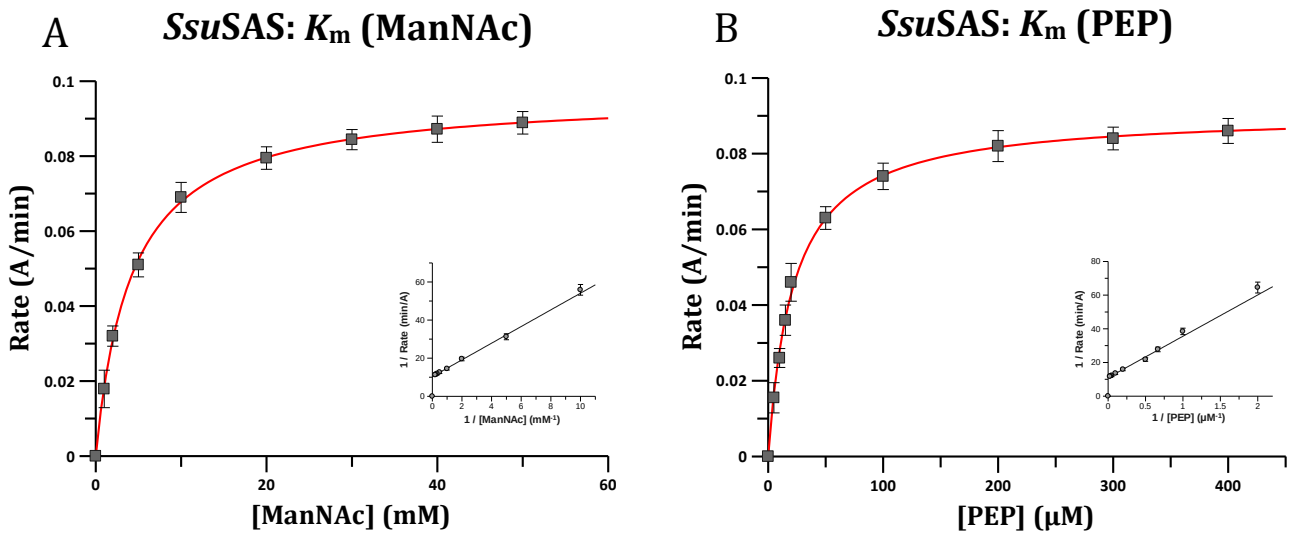


Figure 5.6 Michaelis-Menten curves for *Ssu*SAS with ManNAc and PEP. (A) *Ssu*SAS PEP K_m curve and (B) *Ssu*SAS ManNAc K_m curve. Data points are shown in black and a line of best fit in red. A double reciprocal plot is displayed as an inset for each graph.

	***Nme*NANAS**	***Ssu*SAS**
k_{cat} (s^{-1})	3.1 ± 0.1	0.6 ± 0.02
K_m PEP (µM)	28 ± 3	24 ± 2
K_m ManNAc (mM)	2.9 ± 0.2	4.2 ± 0.1
k_{cat}/K_m ManNAc (s^{-1}/M)	1068 ± 108	142 ± 8

Table 5.2 Kinetic parameters for *Ssu*SAS compared with *Nme*NANAS.

Displayed in Table 5.2, the k_{cat} for *Ssu*SAS was determined to be 1/5 that for wild-type *Nme*NANAS under the same conditions. The K_m (ManNAc) for *Ssu*SAS was determined to be slightly greater than that determined for wild-type *Nme*NANAS at 4.2 ± 0.1 mM and 2.9 ± 0.2 mM respectively. The K_m (PEP) value for *Ssu*SAS was very similar to that determined for *Nme*NANAS at 24 ± 2 µM and 28 ± 3 µM respectively. Unfortunately, k_{cat} and K_m parameters for *Sag*NANAS were not reported by Suryanti *et al* (*105*) and thereby a comparison to the values obtained for *Ssu*SAS could not be made.

5.3.2 pH dependence of activity

In order to determine the optimal pH for activity of *Ssu*SAS, changes in reaction rate were examined using the kinetic assay described in Section 8.5.1 at a variety of different pH values. The buffers used were varied over different pH ranges; acetate was used from pH 4 - 5.5, 2-(*N*-morpholino) ethanesulfonic acid (MES) was used from pH 5.5 – 6.5 and bis-tris propane was used from pH 6.5 – 8.5. Substrates were confirmed to be at saturating levels for each pH analysed. The reaction rate determined at each pH was repeated in triplicate and the average rate plotted graphically (Figure 5.7).

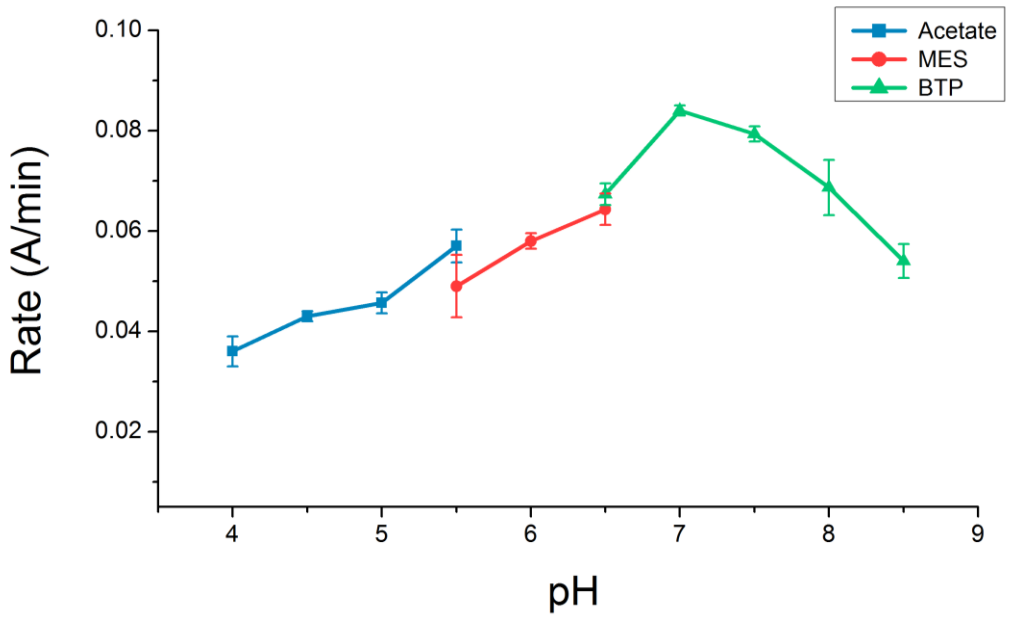


Figure 5.7 pH dependence of *Ssu*SAS activity. Data taken in triplicate and plotted as an average rate (A.min^{-1}). Assays performed at 25 °C with 1mM Mn^{2+}, 50 mM ManNAc and 0.2 mM PEP.

Shown in Figure 5.7, the activity of *Ssu*SAS increased steadily from pH 4 to 7 and gradually dropped away between pH 7.5 to 8.5. Due to the rapid oxidation of Mn^{2+}, data beyond pH 8.5 could not be accurately measured. The different buffers used did not appear to have a major affect on the activity of *Ssu*SAS. From this experiment it was determined that pH 7 is optimal for *Ssu*SAS activity.

5.3.3 Temperature dependence of activity

In order to determine the optimal temperature for activity of *Ssu*SAS, changes in reaction rate were examined using the kinetic assay described in Section 8.5.1 at a variety of different temperatures. Buffers used for each data point were made by standardising to pH 7 while equilibrated at the appropriate temperature. Substrates were confirmed to be at saturating levels for each temperature analysed. The reaction rate determined at each temperature was repeated in triplicate and the average rate was plotted graphically (Figure 5.8).

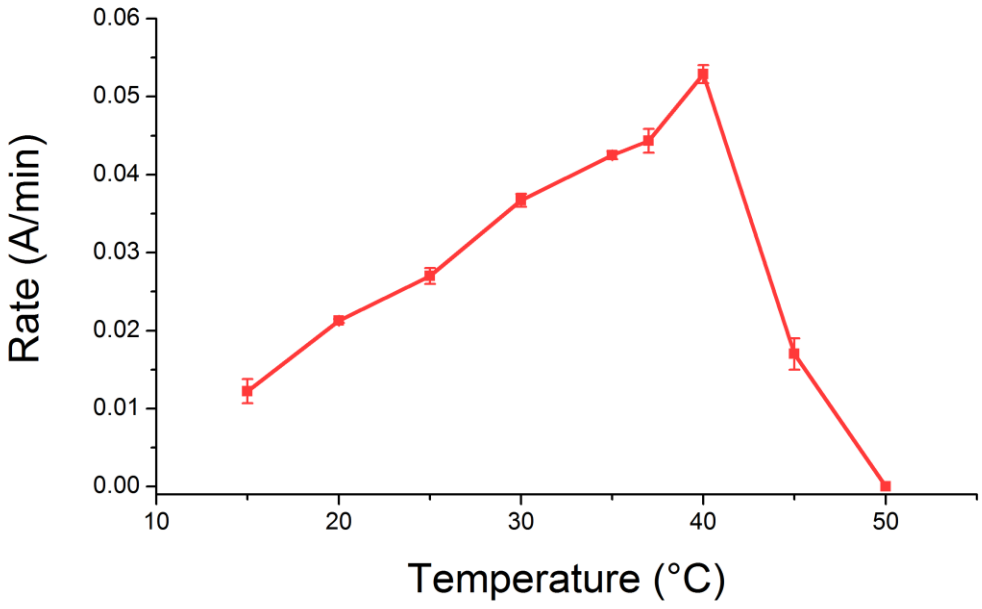


Figure 5.8 Temperature dependence of *Ssu*SAS activity. Data taken in triplicate and plotted as an average rate ($A.min^{-1}$). Assays performed at pH 7 with 1mM Mn^{2+}, 50 mM ManNAc and 0.2 mM PEP.

As displayed in Figure 5.8, the activity of *Ssu*SAS, increased steadily from 15 °C to 40 °C and then dropped rapidly to 50 °C where no activity was observed. From this experiment it was determined that a temperature of around 40 °C is optimal for *Ssu*SAS activity. Given that the host organism for this enzyme, *S. suis* is a pathogen of pigs, the enzyme is likely to operate around 37 °C *in vivo*. The complete loss of activity observed at 50 °C is likely due to denaturation of the enzyme. Although 40 °C was determined to be the optimal temperature for *Ssu*SAS, the enzyme was characterised at 25 °C for comparison of kinetic parameters with *Nme*NANAS.

5.3.4 Metal dependence of activity

The metal dependence of *Ssu*SAS activity was examined using the kinetic assay described in Section 8.5.1 and substituting $MnCl_2$ with a variety of different metal chlorides. Enzyme was equilibrated with metal and the reaction was initiated by the addition of ManNAc. To remove residual activity, metal was also substituted with 1 mM and 10 mM EDTA.

The kinetic effect of different metal ions was analysed using holoenzyme and apoenzyme *Ssu*SAS. Holoenzyme has had access to metal ions from the expression cell line and IMAC chromatography. Apoenzyme was prepared through an additional purification step where enzyme was incubated with 10 mM EDTA for 2 hours at 4 °C after IMAC, to sequester undesired metal ions. EDTA was then separated from the enzyme using SEC. Buffers used for subsequent steps of apoenzyme purification were treated with Chelex® 100 resin (Sigma-Aldrich), and all glassware used was washed with concentrated nitric acid. Each kinetic measurement was determined in triplicate. The percentage relative activity compared to holoenzyme with no metal added was calculated and plotted graphically (Figure 5.9).

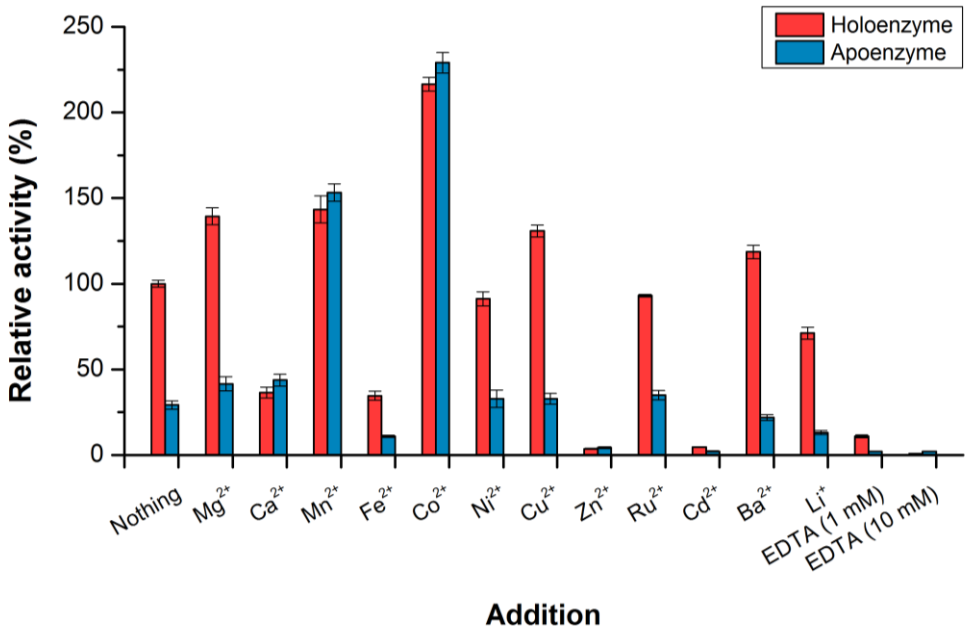


Figure 5.9 Metal dependence of *Ssu*SAS activity. Relative activity was calculated by comparison to holoenzyme with no metal added. Holoenzyme data is shown in blue and apoenzyme data is shown in red. Data was taken in triplicate and averaged prior to calculation of percentage change. Assays performed at 25 °C and pH 7 with 50 mM ManNAc, 0.2 mM PEP. Metals used at 1 mM.

As shown in Figure 5.9, *Ssu*SAS was most strongly activated by Co^{2+} followed by Mn^{2+} for both holoenzyme and apoenzyme. This trend has also been observed previously in *Sag*NANAS (*105*), however, for both *Cje*NANAS and *Nme*NANAS, Mn^{2+} was determined as the most strongly activating metal ion followed by Co^{2+}(*90, 104*). Although Co^{2+} was found to be the most activating metal of *Ssu*SAS, the enzyme was characterised with Mn^{2+} for comparison of kinetic parameters with *Nme*NANAS. Furthermore, the kinetic analysis described in the previous analysis of *Sag*NANAS was determined for holoenzyme without the addition of metal (*105*).

Interestingly, both *Ssu*SAS and *Sag*NANAS (*105*) are activated by Mg^{2+}, as are both *Cje*NANAS and *Nme*NANAS (*90, 104*). The activation of these enzymes by Mg^{2+} indicates that the metal ion plays an important role in substrate binding/phosphate recognition (*173*). As depicted in Figure 5.4, the phosphate group of PEP acts as one of the ligands to the Mn^{2+} ion in the crystal structure of *Nme*NANAS 1XUZ (*112*). Given the similarity of inner and outer shell complexation for Mn^{2+} and Mg^{2+}, it is likely that Mg^{2+} coordinates in a similar fashion depicted for Mn^{2+}.

*Ssu*SAS was also determined to be strongly inhibited by both Zn^{2+} and Cd^{2+} and to a lesser extent inhibited by Fe^{2+}, Ca^{2+}, and then Li^{+} (Figure 5.9). Both Ni^{2+} and Ru^{2+} showed no effect on enzyme activity whilst Cu^{2+} and Ba^{2+} increased the activity of holoenzyme but showed no effect on apoenzyme.

Addition of 1 mM and 10 mM EDTA abolished activity of holoenzyme and apoenzyme revealing the importance of a divalent metal ion for enzyme activity. Displayed in Figure 5.9, both holoenzyme and to a lesser extent apoenzyme had residual NANAS activity. It is likely that this was caused by the enzyme adventitiously scavenging metal ions during the purification process. The addition of EDTA prior to SEC for purification of apoenzyme diminished residual activity to 25% that observed for holoenzyme, suggesting that some metal ion was still present in 'apoenzyme' *Ssu*SAS.

5.3.5 Attempted alternative substrate analysis

As aforementioned, sugar substrate promiscuity has been previously observed with the SAS from *S. agalacticae* (*105*). Activity of *Sag*NANAS with GlucNAc and GalNAc was described as being 15 % and 5 % respectively of the activity measured for enzyme with ManNAc. *Ssu*SAS was calculated to have 81 % sequence identity with *Sag*NANAS and was thus tested for substrate promiscuity with a range of ManNAc analogues (Table 5.3). Unfortunately, no detectable activity could be measured with any of the ManNAc analogues depicted below, even in the presence of 10-fold excess enzyme. Substrates were analysed at 25 °C and 40 °C and at pH 7 and 8 with either 1 mM Mn^{2+}, Mg^{2+} or Co^{2+} however no activity could be observed. Lack of activity equates to a k_{cat} of less than 0.005 s^{-1}.

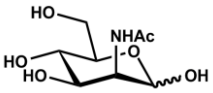
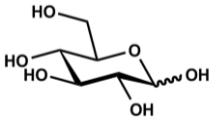
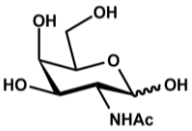
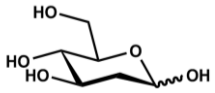
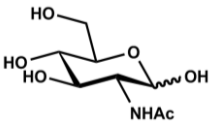
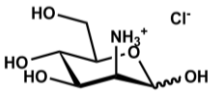
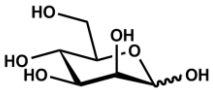
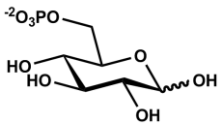
Name	Structure	Name	Structure
***N*-acetyl-D-mannosamine**		**D-glucose**	
***N*-acetyl-D-galactosamine**		**2-deoxy-D-glucose**	
***N*-acetyl-D-glucosamine**		**D-mannosamine**	
D-mannose		***N*-acetyl-D-mannosamine-6-phosphate**	

Table 5.3 Structures of ManNAc and ManNAc analogues assayed with *Ssu*SAS.

5.4 Binding studies of *Ssu*SAS

5.4.1 Isothermal titration calorimetry

Binding constants for Mn^{2+} and PEP were determined for the *Ssu*SAS using the methods described for *Nme*NANAS wild-type (Section 8.5.4). ITC results attained for *Ssu*SAS are compared to previously determined dissociation constants (K_d) for *Nme*NANAS wild-type in Table 5.4 (*114*). The raw experimental binding curves and one-site model fitted ITC data for Mn^{2+} and PEP with *Ssu*SAS are displayed in Figure 5.10.

	***Nme*NANAS (*114*)**	***Ssu*SAS**
K_d Mn^{2+} (μM)	10.6 ± 0.5	3.4 ± 0.2
K_d PEP (μM)	219 ± 11	312 ± 9

Table 5.4 ITC derived binding constants for wild-type *Nme*NANAS and wild-type *Ssu*SAS.

To determine the K_d value for metal with *Ssu*SAS, Mn^{2+} was titrated into metal-free enzyme. The Mn^{2+} affinity of *Ssu*SAS was determined to be approximately 3-fold higher than that previously determined for *Nme*NANAS (Table 5.4). As depicted in Figure 5.10A, the binding of Mn^{2+} is an exothermic event in *Ssu*SAS and is thus an enthalpically favourable process. This is also observed for binding of Mn^{2+} with *Nme*NANAS wild-type (Chapter 2).

Depicted in Table 5.4, the PEP affinity for *Ssu*SAS is slightly lower than that observed for *Nme*NANAS wild-type but is not significantly different. As shown in Figure 5.10B, the binding of PEP is an endothermic event in *Ssu*SAS. This correlates with previous observations for PEP binding with wild-type *Nme*NANAS, which was determined to be an entropically favourable event (Chapters 2 and 3). These results indicate that the *Ssu*SAS Mn^{2+} and PEP binding sites behave similarly to the corresponding sites in *Nme*NANAS. This is unsurprising given that each of the residues involved with metal and PEP binding are fully conserved in both species (Section 5.1.4).

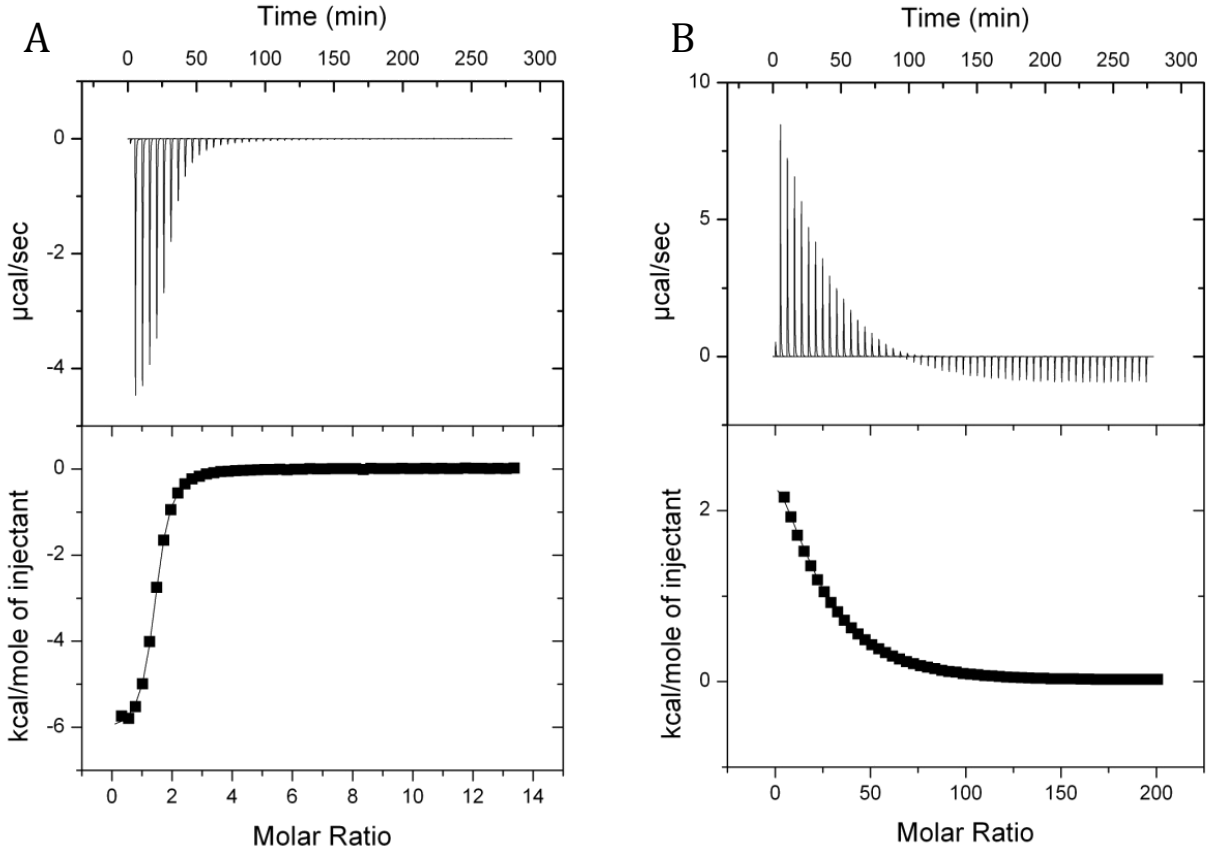


Figure 5.10 Enthalpy changes measured by isothermal titration calorimetry. (A) Mn^{2+} (5 mM) titrated into metal-free *Ssu*SAS (65 µM). (B) PEP (50 mM) titrated into *Ssu*SAS (40 µM) in the presence of Mn^{2+} (1 mM).

5.4.2 Thermal stability with metals

The effect of metal ions on the thermal stability of apoenzyme *Ssu*SAS was assessed by DSF. Melt temperatures of *Ssu*SAS were determined in the presence of a variety of different metal ions using the method outlined in Section 8.5.3. Samples were run in triplicate with 1 mM metal chloride. Controls were run with either no additive or 10 mM EDTA. The melting points determined for each sample are plotted in Figure 5.11 and compiled in Table 5.5.

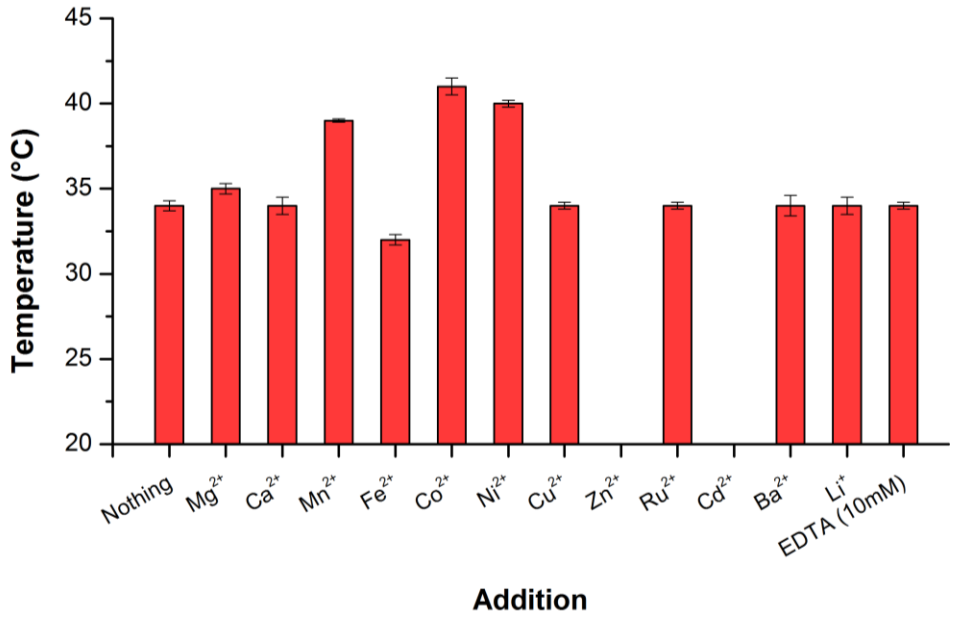


Figure 5.11 Bar graph of thermal melting points for *Ssu*SAS (red) in the presence of metal as determined by DSF. Metal chloride salts were added at a concentration of 1 mM and EDTA at 10 mM.

Addition	*Ssu*SAS (°C)
Nothing	34.0 ± 0.3
Mg^{2+}	34.9 ± 0.3
Ca^{2+}	33.8 ± 0.5
Mn^{2+}	39.2 ± 0.1
Fe^{2+}	31.7 ± 0.3
Co^{2+}	41.3 ± 0.5
Ni^{2+}	40.0 ± 0.2
Cu^{2+}	34.1 ± 0.2
Zn^{2+}	-
Ru^{2+}	33.8 ± 0.2
Cd^{2+}	-
Ba^{2+}	34.2 ± 0.5
Li^{+}	34.1 ± 0.5
EDTA (10 mM)	34.1 ± 0.2

Table 5.5 Thermal melting points for *Ssu*SAS in the presence of metal as determined by DSF. Metal chloride salts were added at a concentration of 1 mM and EDTA at 10 mM. – indicates protein denaturation upon addition of metal.

Apoenzyme *Ssu*SAS was determined to melt at 34.0 ± 0.3 °C (Figure 5.11). *Apo*-*Ssu*SAS is far less stable than *apo*-*Nme*NANAS which melts at 47.0 ± 0.2 °C (Chapter 2). The addition of 1 mM Mn^{2+} increased the melting point of *Ssu*SAS, by approximately 5 °C, similar to what is observed with *Nme*NANAS. As shown in Figure 5.11, Co^{2+}, Ni^{2+} and Mn^{2+} each increased the melting temperature of *Ssu*SAS. While both Co^{2+} and Mn^{2+} activate *Ssu*SAS, Ni^{2+} was shown to have little effect on activity (Figure 5.9). Of the metals analysed, *Ssu*SAS was most thermally stabilised by Co^{2+} and Mn^{2+} which produced similar melting temperatures of 41.3 ± 0.5 °C and 39.2 ± 0.1 °C respectively. Described in Section 5.3.4 both Co^{2+} and Mn^{2+} were also the most activating of the metals analysed with *Ssu*SAS.

Mg^{2+}, Ca^{2+}, Cu^{2+}, Ru^{2+}, Ba^{2+}, Li^{+}, and 10 mM EDTA had little to no effect on the melting point of *Ssu*SAS, whereas Fe^{2+}, which was shown to inhibit *Ssu*SAS (Figure 5.9), lowered the melting point of *apo-Ssu*SAS to 31.7 ± 0.3 °C. The observation that EDTA did not influence the thermal stability of the enzyme likely indicates the successful removal of trace metal ions during purification. Interestingly, both Zn^{2+} and Cd^{2+}, which were shown to inhibit *Ssu*SAS activity completely (Figure 5.9), showed high fluorescence of SYPRO orange dye from the onset of the temperature melt experiment, indicating that the enzyme had been unfolded upon addition of the metal. It is likely that the loss of activity observed with 1 mM Zn^{2+} and Cd^{2+} was caused by denaturation of the protein.

5.4.3 Thermal stability with substrates and analogues

The effect of natural substrates and ManNAc analogues on the thermal stability of apoenzyme *Ssu*SAS was assessed by DSF using the method outlined in Section 8.5.3. Combinations of metal, PEP and ManNAc/ManNAc analogue were used for analysis. The effects of the substrates and analogues upon the thermal stability of *Ssu*SAS were analysed in combination with either Mn^{2+} or Co^{2+}. The melting points determined for each sample are displayed in Figure 5.12 and Table 5.6.

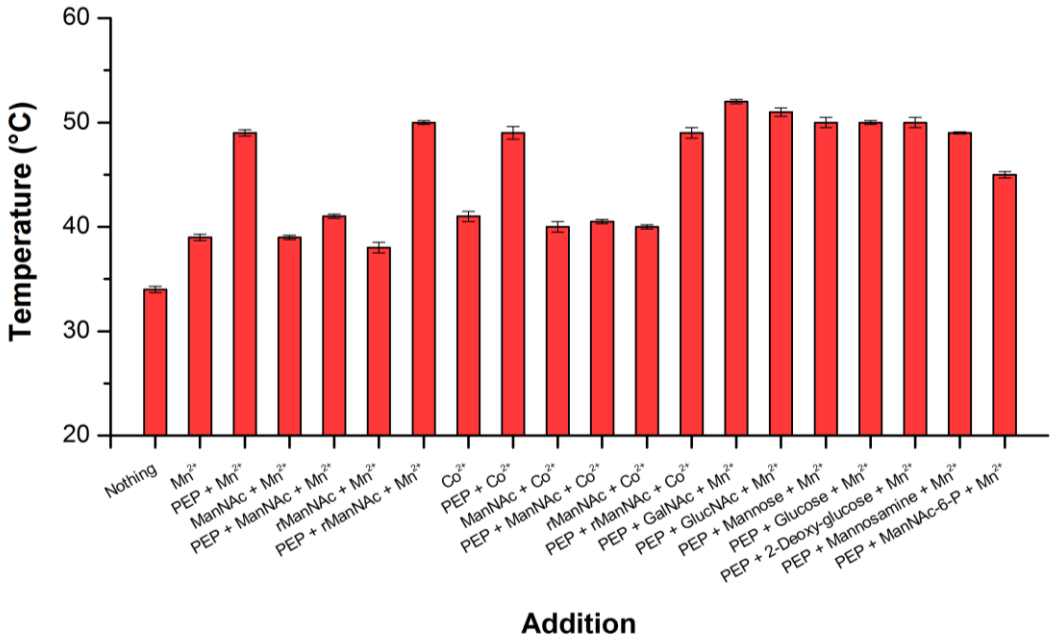


Figure 5.12 Bar graph of thermal melting points for *Ssu*SAS (red) in the presence of ligands as determined by DSF. Metal chloride salts were added at 1 mM, PEP at 1 mM and ManNAc/ManNAc Analogues at 10 mM with metal and at 1 mM when added in combination with PEP and metal.

Addition	*Ssu*SAS (°C)
Nothing	34.0 ± 0.3
Mn^{2+}	39.2 ± 0.1
PEP + Mn^{2+}	49.1 ± 0.3
ManNAc + Mn^{2+}	38.9 ± 0.2
PEP + ManNAc + Mn^{2+}	41.0 ± 0.2
rManNAc + Mn^{2+}	38.0 ± 0.5
PEP + rManNAc + Mn^{2+}	50.1 ± 0.2
Co^{2+}	41.3 ± 0.5
PEP + Co^{2+}	49.2 ± 0.6
ManNAc + Co^{2+}	40.1 ± 0.5
PEP + ManNAc + Co^{2+}	40.5 ± 0.2
rManNAc + Co^{2+}	40.1 ± 0.2
PEP + rManNAc + Co^{2+}	49.0 ± 0.5
PEP + GalNAc + Mn^{2+}	52.3 ± 0.2
PEP + GlucNAc + Mn^{2+}	51.4 ± 0.4
PEP + Mannose + Mn^{2+}	50.1 ± 0.5
PEP + Glucose + Mn^{2+}	49.8 ± 0.2
PEP + 2-Deoxy-glucose + Mn^{2+}	50.0 ± 0.5
PEP + Mannosamine + Mn^{2+}	49.2 ± 0.1
PEP + ManNAc-6-P + Mn^{2+}	45.4 ± 0.3

Table 5.6 Thermal melting points for *Ssu*SAS in the presence of ligands as determined by DSF. Metal chloride salts were added at 1 mM, PEP at 1 mM and ManNAc/ManNAc Analogues at 10 mM with metal and at 1 mM when added in combination with PEP and metal.

The thermal melt profile of *Ssu*SAS is very similar to that observed for *Nme*NANAS (Chapter 2). Similar to *Nme*NANAS, an enormous thermal shift is observed for *Ssu*SAS in the presence of both Mn^{2+} and PEP. The combined addition of 1 mM Mn^{2+} and 1 mM PEP to *Ssu*SAS increased the melting point of apoenzyme from 34.0 ± 0.3 °C to 49.1 ± 0.3 °C. The observation of such a dramatic temperature change induced by the presence of both Mn^{2+} and PEP, likely indicates a significant conformational change that stabilises the enzyme.

As depicted in Figure 5.12 and Table 5.6, *Ssu*SAS in the presence of Mn^{2+} and ManNAc/rManNAc mirrors the melting temperature of *Ssu*SAS in the presence of Mn^{2+}. This trend is observed with each of the wild-type SASs examined in this thesis. As commented in Chapters 2 and 3, *Nme*NANAS is unable to bind ManNAc without PEP binding first. Given the high conservation of active site residues between *Nme*NANAS and *Ssu*SAS (described in Section 5.1.4), it is likely that a similar mechanism is employed by *Ssu*SAS.

When in combination with Mn^{2+}, PEP and ManNAc, the melting temperature of *Ssu*SAS is very similar to that observed for *Ssu*SAS with just Mn^{2+}. *Ssu*SAS is able to catalyse the formation of NANA from PEP and ManNAc (Section 5.3). Since the aforementioned ligands are in equimolar proportions, *Ssu*SAS will likely be in the presence of Mn^{2+} and NANA. It is unlikely that *Ssu*SAS binds to or is stabilised by NANA to ensure favourable product release. Furthermore, no product inhibition was observed in kinetic analysis of *Ssu*SAS with ManNAc.

The melting temperature of *Ssu*SAS in the presence of Mn^{2+}, PEP and rManNAc is very similar to that of *Ssu*SAS in the presence of Mn^{2+} and PEP (Figure 5.12 and Table 5.6). Since the aldehyde component of rManNAc is removed, *Ssu*SAS can no longer utilise rManNAc as a substrate. The thermal shift observed is thus likely caused by unreacted PEP still present in the active site. This is also observed for *Nme*NANAS under the same conditions.

The thermal melting profile of *Ssu*SAS was primarily analysed using Mn^{2+} in order to compare DSF results with each of the wild-type enzymes mentioned in this thesis. Since Co^{2+} was determined to be the most activating metal for *Ssu*SAS (Section 5.3.4) and most thermally stabilising (Section 5.4.2), DSF of *Ssu*SAS with substrates was also carried out in the presence of Co^{2+}. As observed in Figure 5.12, the thermal melt profile of *Ssu*SAS and substrates is comparable in the presence of either metal.

As outlined in Section 5.3.5, none of the ManNAc analogues assayed with *Ssu*SAS showed any activity including both GalNAc and GlucNAc which had been previously indicated as alternative substrates of *Sag*NANAS (*105*). No loss of PEP was observed for the ManNAc analogues assayed with *Ssu*SAS even with the addition of 10-fold excess enzyme, indicating that *Ssu*SAS was unable to catalyse the aldol condensation between PEP and the alternative sugar compounds.

Shown in Figure 5.12, combinations of PEP and Mn^{2+} with ManNAc analogues did not significantly affect the thermal stability of *Ssu*SAS when compared to the melting point of *Ssu*SAS with only PEP and Mn^{2+} (49.1 ± 0.3 °C). Both GalNAc and GlucNAc did increase the melting temperate of *Ssu*SAS with metal and PEP from 49.1 ± 0.3 °C to 52.3 ± 0.2 °C and 51.4 ± 0.4 °C respectively. This may indicate that although the enzyme is unable to react with either GalNAc or GlucNAc, these compounds are still able to bind to the enzyme. Interestingly, ManNAc-6-P (the natural substrate of the SAS from *H. sapiens*) lowered the melting temperature to 45.4 ± 0.3 °C when in combination with PEP and Mn^{2+}.

5.5 Physical characterisation of *Ssu*SAS

5.5.1 Circular dichroism

CD was used to assess and compare the secondary structure of *Ssu*SAS with *Nme*NANAS (Figure 5.13). The K2D3 server (*142*) was used to estimate the secondary structure composition from CD spectra. Shown in Table 5.7, estimated secondary structure is compared to values derived from the crystal structure of *Nme*NANAS (PDB code: 1XUZ) and the crystal structure of *Ssu*SAS outlined in Section 5.6.

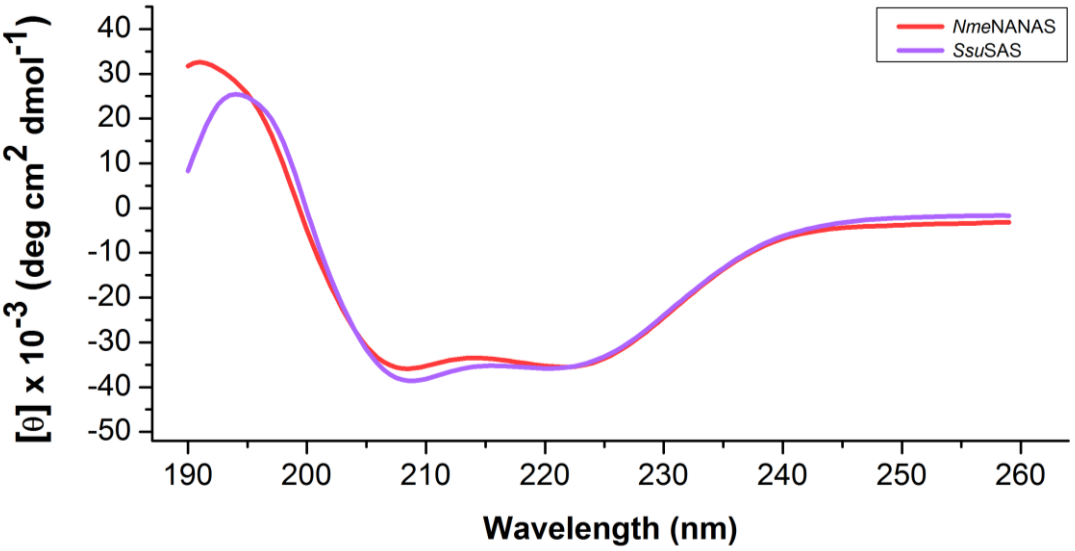


Figure 5.13 CD spectra for *Nme*NANAS (red) and *Ssu*SAS (purple).

	***Nme*NANAS**	***Nme*NANAS (structure)**	***Ssu*SAS**	***Ssu*SAS (structure)**
α-helix	59.9 %	50.3 %	57.9 %	40.1 %
β-sheet	12.3 %	15.5 %	16.7 %	17.2 %
other	27.8 %	34.2 %	25.4 %	42.7 %

Table 5.7 Predictions of secondary structure composition for *Nme*NANAS and *Ssu*SAS calculated from CD data using K2D3 server (*142*). Secondary structure composition of *Nme*NANAS structure calculated from crystallographic coordinates 1XUZ and secondary structure composition of *Ssu*SAS structure calculated from crystal structure outlined in Section 5.6.

As depicted in Figure 5.13, *Ssu*SAS is properly folded in solution and has a similar CD profile to *Nme*NANAS. Shown in Table 5.7, the levels of α-helix and β-sheet for the CD profiles of *Nme*NANAS and *Ssu*SAS are comparable to the secondary strucutre ratios calculated from *Nme*NANAS crystal structure 1XUZ. Interestingly, the predicted secondary structure ratios calculated from the *Ssu*SAS crystal structure are fairly different to those derived from both the CD profile of *Ssu*SAS and the secondary structure ratio calculated from *Nme*NANAS 1XUZ. This is likely due to the secondary strucuture in the *Ssu*SAS crystal strucutre being less defined when compared with *Nme*NANAS 1XUZ.

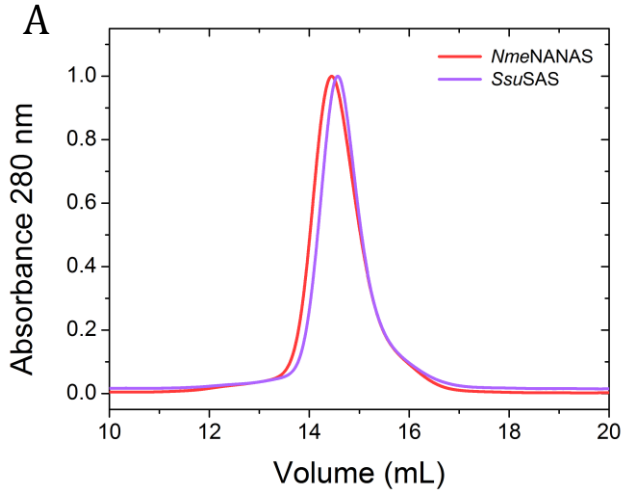
5.5.2 Mass spectrometry

The molecular mass of His_6-tagged *Ssu*SAS was analysed using MS. A theoretical mass of 41,160 Da was calculated for *Ssu*SAS from the peptide sequence using ProtParam (*141*). The experimental mass determined by MS was 41,159 Da which was almost identical to the aforementioned calculated mass.

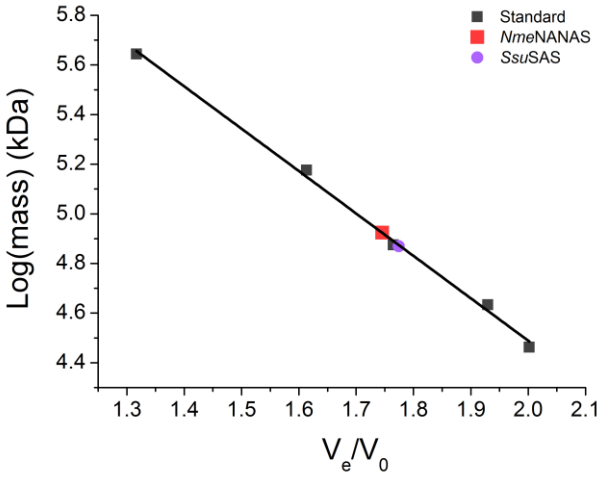
5.5.3 Analytical size exclusion

The quaternary state of *Ssu*SAS was investigated via analytical SEC using the method outlined in Section 8.5.6. A calibration curve was generated using protein standards of known molecular weight. The elution volume (V_e) of protein standards relative to the column void volume (V_0) was plotted against the log of the protein standards molecular mass. A linear fit of the analytical SEC calibration data was determined to have an R^2 value of 0.987. The calibration curve was used to extrapolate the molecular mass of each of *Ssu*SAS as shown in Figure 5.14.

A



B



C

	***Nme*NANAS**	***Ssu*SAS**
V_e/V_0	1.74	1.77
Estimated mass (kDa)	81	75
Theoretical mass of dimer (kDa)	77	75

Figure 5.14 (A) Normalised analytical SEC chromatograms for *Nme*NANAS (red) and *Ssu*SAS (purple). (B) Analytical SEC calibration curve. Protein standards are marked in grey, *Nme*NANAS in red and *Ssu*SAS in purple. A line of best fit is plotted in black. (C) Table of estimated and theoretical masses for *Nme*NANAS and *Ssu*SAS.

Depicted in Figure 5.14A, *Ssu*SAS elutes as a single species. The molecular weight of *Ssu*SAS derived from the standard curve corresponds to 75 kDa (Figure 5.14B). The molecular weight of *Ssu*SAS determined from analytical SEC is almost identical to the theoretical mass for dimeric *Ssu*SAS. The observation that *Ssu*SAS is dimeric correlates with the results obtained from analytical SEC for each of the wild-type SASs examined in this thesis.

5.6 Structure of *Ssu*SAS

5.6.1 Crystallisation

Purified *Ssu*SAS was crystallised using a condition obtained from a Morpheus™ crystallisation screen (Molecular Dimensions). Monoclinic crystals of *apo-Ssu*SAS were formed using the hanging drop crystallisation technique in the presence of 0.09 M $NaNO_3$/Na_2HPO_4/$(NH_4)_2SO_4$, 0.1 M Imidazole/MES and 37.5% racemic MPD/PEG 1000/PEG 03350 at pH 6.5 (Figure 5.15). Data sets were collected at the Australian Synchrotron using the MX2 beamline.

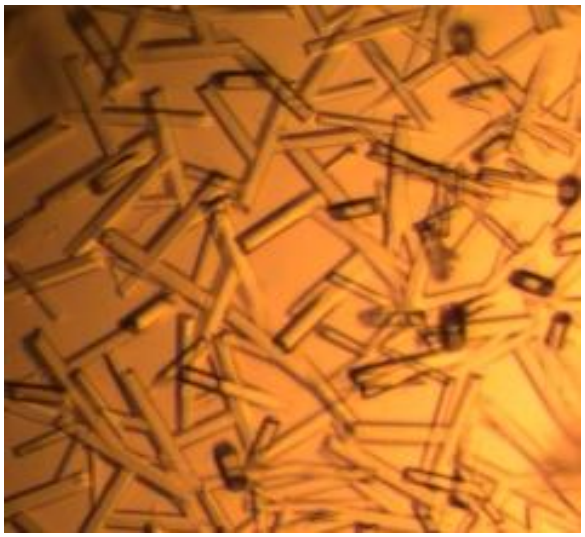


Figure 5.15 Image of rod shaped *Ssu*SAS crystals viewed through a Leica light microscope.

5.6.2 Crystal structure of *Ssu*SAS

A crystal structure of *Ssu*SAS was successfully solved to a resolution of 2.30 Å. The structure of *Ssu*SAS was solved via molecular replacement using a model generated from the wild-type *Nme*NANAS crystal structure 1XUZ (*112*), which had the linker, AFPL domain, ligands and non-conserved side-chains removed. The AFPL domains and linker domains for both monomers of *Ssu*SAS were manually built during refinement. The methods for molecular replacement and refinement of the *Ssu*SAS structure are outlined in Section 8.5.12. Due to anisotropy in the diffraction data, the resolution cut-off for the *Ssu*SAS structure was made at 2.30 Å. The final refinement statistics for the structure of *Ssu*SAS are given in Table 5.8 (*Ssu*SAS structure currently not deposited).

	*Ssu*SAS
A. *Data collection*	
Crystal system; space group	$P12_11$
Unit cell parameters (Å)	82.03, 58.19, 90.49,
a, b, c, α, β, γ	90.00, 108.04, 90.00
Resolution range (outer shell) (Å)	48.20-2.30 (2.38-2.30)
Measurements	137,121
Unique reflections	36,417
Redundancy	3.4
Completeness (outer shell) (%)	99.4 (99.7)
I/σ(outer shell) (I)	6.7 (2.6)
R_{merge} (outer shell)	0.128 (0.502)
Wilson B-value (Å^2)	13.824
B. *Refinement*	
Resolution (Å)	2.30
R_{cryst}	0.1924
R_{free}	0.2545
Chain length	337
Observed number of residues	674
Water molecules	231
Other	4
Mean B (Å^2)	
Protein	13.8
Water	26.4
Other	44.3
r.m.s.d. from target values	
Bond lengths (Å)	0.0165
Bond angles (°)	1.8965
Dihedral angles (°)	0.1048
Ramachandran	
Most favored (%)	97.3
Allowed (%)	2.7
Generously allowed (%)	0
Disallowed (%)	0

Table 5.8 Crystal parameters, data collection and refinement statistics for crystal structure of *Ssu*SAS.

Displayed in Figure 5.16, the *Ssu*SAS crystal structure is a domain swapped homodimer similar to the previously determined structures of wild-type *Nme*NANAS (*112, 113*). Each monomer of *Ssu*SAS comprises an N-terminal $(\beta/\alpha)_8$ barrel catalytic domain with a short linker sequence to a C-terminal AFPL domain. The dimeric form of *Ssu*SAS is unsurprising, given the aforementioned results obtained from analytical SEC, SAXS and AUC as well as the comparable sequence homology to *Nme*NANAS.

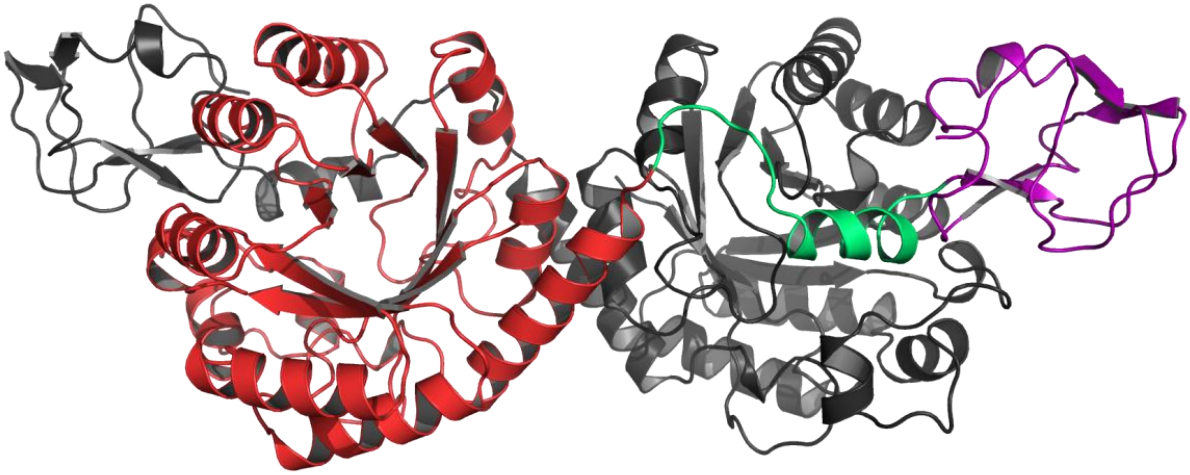


Figure 5.16 Crystal structure of *Ssu*SAS domain-swapped homodimer. Chain A is coloured by domain and Chain B is in grey. The Chain A catalytic domain is in red, linker domain in green and AFPL domain in purple.

Although holistically similar, a comparison of the *Ssu*SAS crystal structure with previously obtained structures of *Nme*NANAS reveals several key differences. Firstly, the unit cell dimensions differ between *Ssu*SAS and *Nme*NANAS, as do the space groups which are $P12_11$ and $P2_12_12$ respectively. For the structures of *Nme*NANAS, the dimeric enzyme is generated through symmetry, whereas for the structure of *Ssu*SAS the full dimer is observed within the unit cell.

Conformational differences between the crystal structures of *Ssu*SAS and *Nme*NANAS are highlighted through structural alignments. RMSD values for alignments between the *Ssu*SAS crystal structure and *Nme*NANAS 1XUZ are compiled in Table 5.9.

Alignment	RMSD (Atoms in alignment)
*Ssu*SAS dimer to *Nme*NANAS dimer	1.961 Å (524 to 524 atoms)
*Ssu*SAS chain A to *Ssu*SAS chain B	0.146 Å (317 to 317 atoms)
*Ssu*SAS chain A to *Nme*NANAS monomer	1.704 Å (254 to 254 atoms)
*Ssu*SAS chain B to *Nme*NANAS monomer	1.698 Å (254 to 254 atoms)
*Ssu*SAS chain A catalytic domain to *Ssu*SAS chain B catalytic domain	0.130 Å (236 to 236 atoms)
*Ssu*SAS chain A catalytic domain to *Nme*NANAS catalytic domain	1.126 Å (215 to 215 atoms)
*Ssu*SAS chain B catalytic domain to *Nme*NANAS catalytic domain	1.104 Å (215 to 215 atoms)
*Ssu*SAS chain A AFPL domain to *Ssu*SAS chain B AFPL domain	0.124 Å (55 to 55 atoms)
*Ssu*SAS chain A AFPL domain to *Nme*NANAS monomer AFPL domain	1.658 Å (31 to 31 atoms)
*Ssu*SAS chain B AFPL domain to *Nme*NANAS monomer AFPL domain	1.168 Å (31 to 31 atoms)

Table 5.9 RMSD values for structural alignments between the crystal structures of *Ssu*SAS and *Nme*NANAS. The crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*).

Dimer and monomer structural alignments of the *Ssu*SAS crystal structure with *Nme*NANAS 1XUZ are displayed in Figure 5.17. The RMSD between the dimer of *Ssu*SAS and dimer of 1XUZ was calculated to be 1.961 Å. Denoted in Table 5.9, Chains A and B of the *Ssu*SAS structure are conformationally similar and have an RMSD value of 0.146 Å. Depicted in Figure 5.18B, *Ssu*SAS Chain A and Chain B, have RMSD values of 1.704 Å and 1.698 Å respectively, when aligned to the monomer of *Nme*NANAS 1XUZ.

In Figure 5.17, dimer and monomer alignments for the substrate-free structure of *Ssu*SAS with the substrate-bound structure of *Nme*NANAS (1XUZ) reveal that the AFPL domains of *Ssu*SAS are positioned closer to the catalytic domains. This degree of domain movement has been observed in MD simulations of the *Nme*NANAS wild-type structure (Chapter 3).

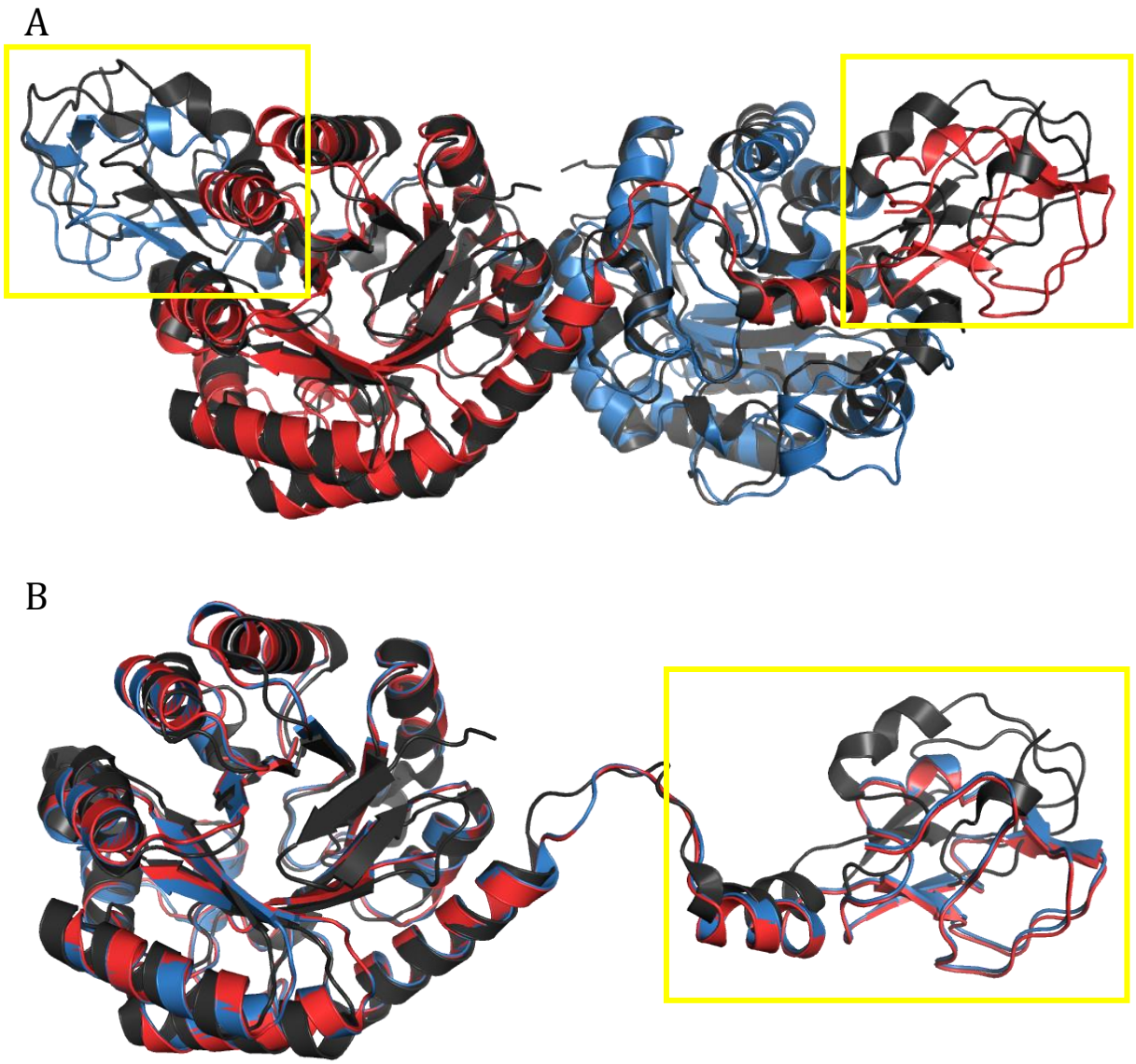
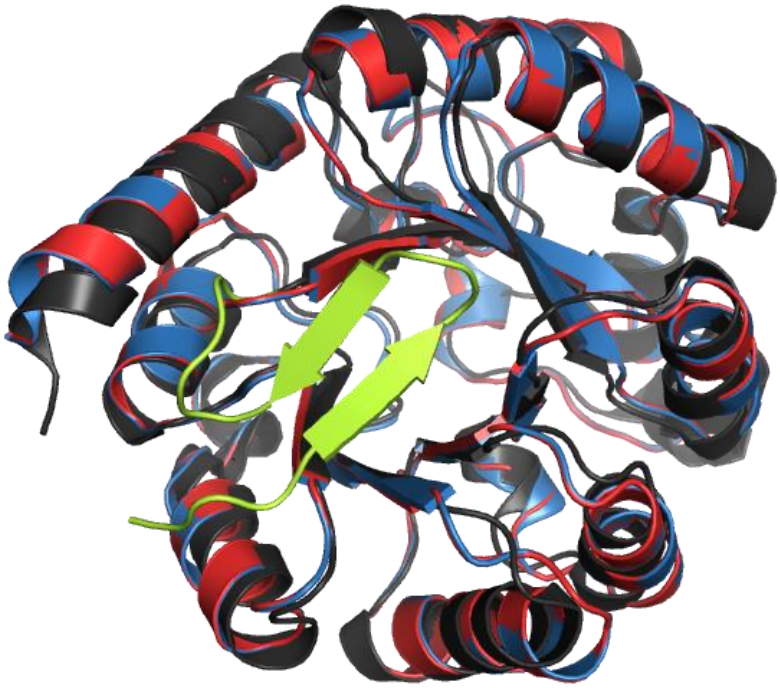


Figure 5.17 *Ssu*SAS and *Nme*NANAS structural alignments. (A) Alignment of *Ssu*SAS and *Nme*NANAS dimers. (B) Alignment of *Ssu*SAS and *Nme*NANAS monomers. *Nme*NANAS is coloured grey, Chain A of the *Ssu*SAS structure is in red and Chain B is in blue. The crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*). The structural differences are highlighted with yellow boxes.

As discussed in Chapter 3, a conformational change in *Nme*NANAS is observed upon liganding of PEP which pre-organises the ManNAc binding site. Evidence obtained from SAXS, ITC and DSF with PEP indicate that a similar mechanism is employed by *Ssu*SAS. It is possible that the observed changes in positioning of the AFPL domains in the *Ssu*SAS structure have resulted from the absence of ligand. Unfortunately, a structural comparison of liganded and ligand-free *Ssu*SAS could not be made, as a liganded crystal structure of *Ssu*SAS was not obtained despite extensive crystallisation trials. Ligand soaking, co-crystallisation and crystal seeding of *Ssu*SAS were attempted with several conditions, however, crystals of appropriate diffraction quality were not afforded.

The conformation of the core N-terminal $(\beta/\alpha)_8$ barrel is mostly retained in both *Ssu*SAS and *Nme*NANAS (Table 5.9), however, there are two key structural differences observed. Displayed in Figure 5.18A, an N-terminal β hairpin loop comprising approximately 20 amino acids is observed in the *Nme*NANAS structure which is not present in *Ssu*SAS. The sequence of this loop is not conserved, as its composition and length varies within the SAS family. As depicted by sequence alignment with *Nme*NANAS (Figure 5.18B), this region is completely missing from the *Ssu*SAS sequence.

A



B

```
SsuSAS/1-338      1 --------------------MVYIIAEIGCNHNGDVHLARKMV 23
NmeNANAS/1-349    1 MQNNNEFKIGNRSVGYNHEPLIICEIGINHEGSLKTAFEMV 41
```

Figure 5.18 *Ssu*SAS and *Nme*NANAS catalytic domain alignments. (A) Structural alignment of *Ssu*SAS and *Nme*NANAS catalytic domains. *Nme*NANAS is coloured grey, Chain A of the *Ssu*SAS structure is in red and Chain B is in blue. The *Nme*NANAS β-hairpin loop is highlighted in yellow. The crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*). (B) Sequence alignment of N-terminal residues from *Ssu*SAS and *Nme*NANAS. Conserved residues are highlighted in blue, missing residues are depicted with dashes.

The second difference between the catalytic domains of *Ssu*SAS and *Nme*NANAS is the β2-α2 loop. The β2-α2 loop closes over the catalytic site and contributes residue ^{Nme}Asn74 in the *Nme*NANAS structure 1XUZ, which forms a hydrogen-bond to the C-6 hydroxyl of ManNAc (Figure 5.19A). Shown in Figure 5.19B, this residue is not conserved in *Ssu*SAS. Instead the *Ssu*SAS β2-α2 loop has a similar sequence to members of the LegS clade of SASs (Figure 5.3). Displayed in Figure 5.19A, the *Ssu*SAS β2-α2 loop forms an α-helix and thus differs in secondary structure composition when compared with the corresponding region from *Nme*NANAS 1XUZ.

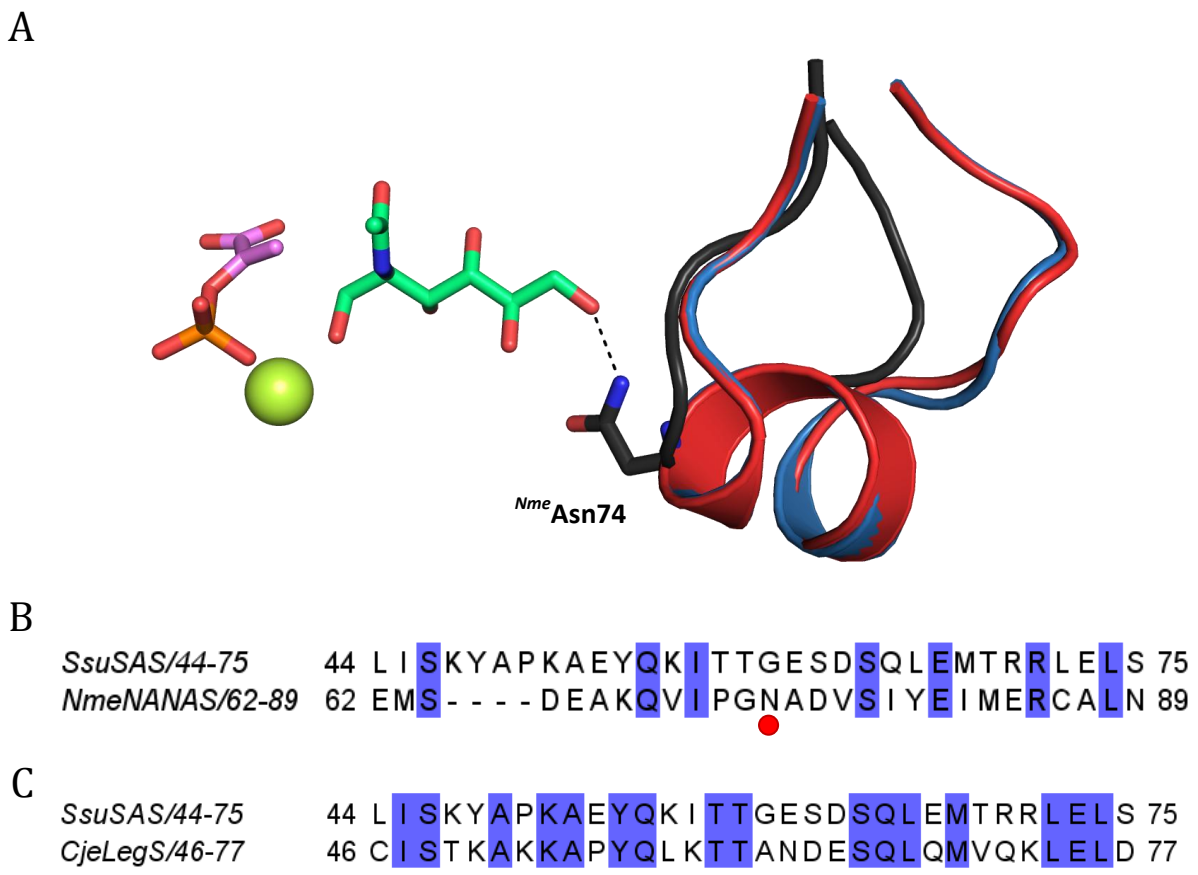


Figure 5.19 *Ssu*SAS and *Nme*NANAS β2-α2 loops. (A) Position of β2-α2 loops from a structural alignment of *Ssu*SAS and *Nme*NANAS catalytic domains. *Nme*NANAS is coloured grey, Chain A of the *Ssu*SAS structure is in red and Chain B is in blue. Ligands are included from the *Nme*NANAS structure with Mn^{2+} in yellow, PEP in pink and rManNAc in green. H-bonds are depicted with black dashes. Crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*). (B) Sequence alignment of β2-α2 loops from *Ssu*SAS and *Nme*NANAS. Conserved residues are highlighted in blue, missing residues are depicted with dashes. ^{Nme}Asn74 is marked with a red circle. (B) Sequence alignment of β2-α2 loops from *Ssu*SAS and *Cje*LegS. Conserved residues are highlighted in blue.

A sequence alignment of the β2-α2 loop from *Ssu*SAS and *Nme*NANAS (Figure 5.19B) reveals that there is significant sequence variation between the two enzymes at this position. Notably, the *Ssu*SAS β2-α2 loop is extended by 4 amino acid residues. Multiple sequence alignments of bacterial SAS enzymes have indicated that this loop is likely to have a major influence on substrate selectivity (Chapter 4). Displayed in Figure 5.20C, the β2-α2 loop from *Ssu*SAS is of identical length and has far greater sequence identity to the β2-α2 loop from the LegS of *C. jejuni* (50 %) than to the β2-α2 loop from *Nme*NANAS (22 %). This observation concurs with the phylogenetic analyses detailed in Sections 5.1.3 and 5.1.4 which also indicate that *Ssu*SAS is a member of the LegS clade.

In addition, the β2-α2 loop has been identified in MD simulations of *Nme*NANAS as a region of significant flexibility when in a ligand-free state (Chapter 3). As previously mentioned, the structure of *Ssu*SAS was attained without substrate present. The combined absence of ligand and flexibility of the β2-α2 loop may indicate that the crystallised conformation of this region in the *Ssu*SAS structure may not be observed when Mn^{2+}, PEP and ManNAc are bound.

Similar to *Nme*NANAS, the N-terminal sequence of *Ssu*SAS has an analogous fold to Type III antifreeze proteins. The RMSD values for alignments of the *Nme*NANAS AFPL domain with Chain A and Chain B AFPL domains from *Ssu*SAS are 1.658 Å and 1.168 Å respectively. Illustrated in Figure 5.20, the AFPL domain of *Ssu*SAS retains most of the secondary structural elements observed in *Nme*NANAS 1XUZ, but is slightly less structured in some regions. The C-terminus of the AFPL domain is extended in *SsuSAS* and does not share the partial helix that is formed by the last five residues in the *Nme*NANAS structure. Instead, the last few C-terminal residues of the *SsuSAS* structure trail away in the opposite direction to that of the *Nme*NANAS structure (Figure 5.20).

^{Nme}Arg314 is an active site residue, contributed by the opposing monomer in *Nme*NANAS which is shown to interact with the *N*-acetyl group of rManNAc in the 1XUZ structure (Figure 5.20). Discussed in Chapter 2, ^{Nme}Arg314 is an essential catalytic residue and is responsible for steering ManNAc into a reactive position. ^{Nme}Arg314 and the corresponding residue in *Ssu*SAS (^{Ssu}Arg301) are in very similar positions relative to the AFPL domains of their respective structures (Figure 5.20).

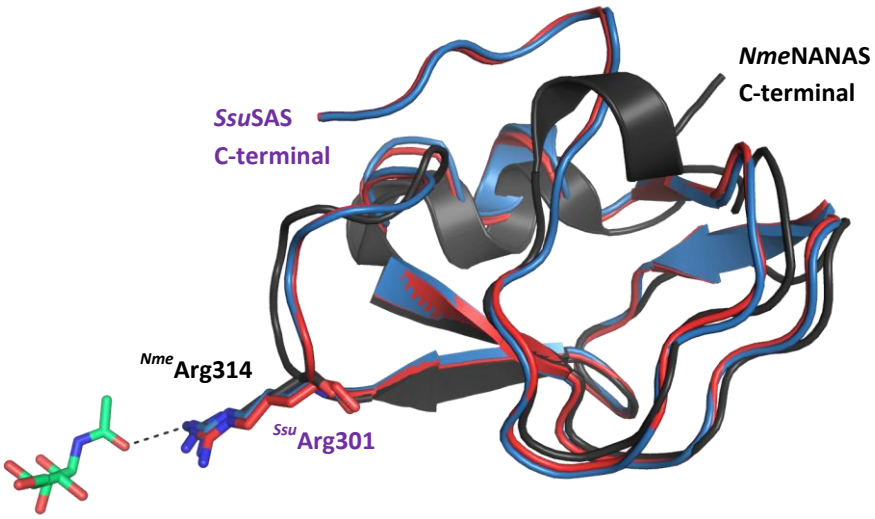


Figure 5.20 Structural alignments of *Nme*NANAS and *Ssu*SAS AFPL domains. *Nme*NANAS is coloured grey, Chain A of the *Ssu*SAS structure is in red and Chain B is in blue. rManNAc (green) is included from the *Nme*NANAS structure. H-bonds are depicted with black dashes. Crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*).

As aforementioned, *Ssu*SAS was crystallised without Mn^{2+}, PEP or rManNAc bound, however, a phosphate ion and glycerol molecule were observed in both chains. Displayed in Figure 5.21, the phosphate ion makes polar contacts with ^{Ssu}Lys115, ^{Ssu}Ser118, ^{Ssu}Ser140 and ^{Ssu}Ser200 of *Ssu*SAS. The corresponding residues in the *Nme*NANAS structure 1XUZ form similar contacts with the phosphate moiety of PEP (Figure 5.22). The glycerol molecule observed in the active site of *Ssu*SAS forms interactions with ^{Ssu}Glu55, ^{Ssu}Arg301 and two water molecules. The most probable source of the phosphate ion is from Na_2HPO_4 which, as mentioned in Section 5.6.1, was a component in the crystallisation conditions for *Ssu*SAS. A glycerol mixture was used as a cryoprotectant for *Ssu*SAS crystals which is likely to have entered the active site, an occurrence which is well documented in the literature (*174-176*).

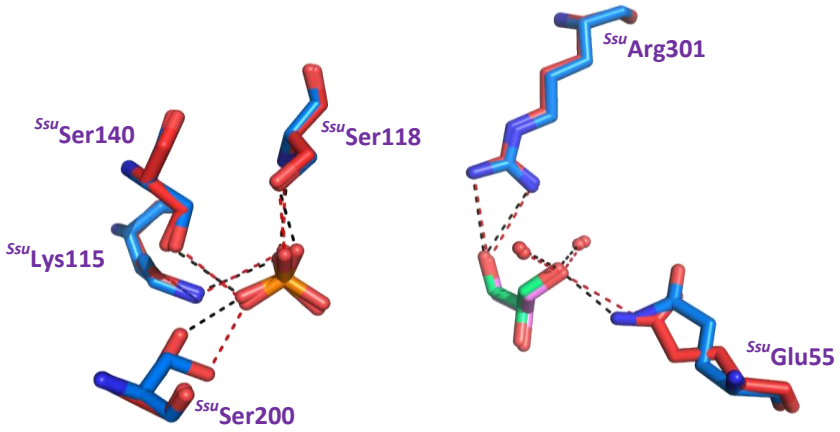


Figure 5.21 Ligand interactions observed in the *Ssu*SAS crystal structure. Chain A of *Ssu*SAS (red) is aligned with Chain B (blue). Bound glycerol from Chain A is in green and from Chain B in pink. Polar contacts to ligands in Chain A are depicted with black dashes and in Chain B with red dashes. Waters are displayed as red spheres.

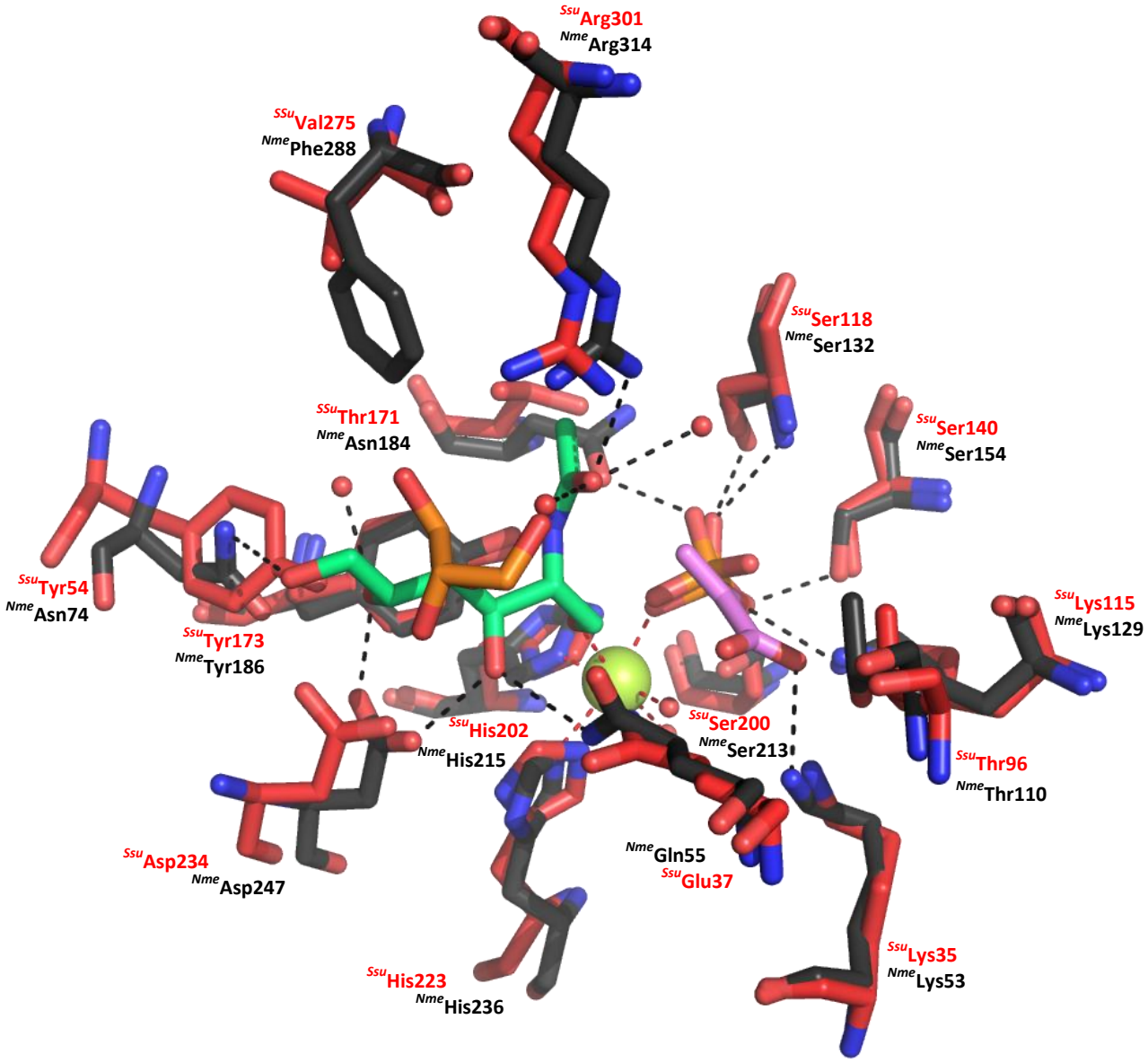


Figure 5.22 Overlay of *Nme*NANAS and *Ssu*SAS active site residues. Residues from *Nme*NANAS structure are coloured grey and Chain A of *Ssu*SAS in red. Ligands are included from the *Nme*NANAS structure with Mn^{2+} in yellow, PEP in pink, rManNAc in green and waters as red spheres. Glycerol is included from Chain A of the *Ssu*SAS structure in orange. Polar contacts are depicted with black dashes and metal-ligand interactions with red dashes. Crystallographic coordinates for the *Nme*NANAS structure are 1XUZ (*112*).

The positioning of the *Ssu*SAS active site residues is analogous to what is observed in the crystal structure of *Nme*NANAS 1XUZ (Figure 5.22). The residues involved with metal coordination (^{Ssu}His202/^{Nme}His215 and *Ssu*His223/*Nme*His236) are located similarly in both 1XUZ and the structure of *Ssu*SAS even though *Ssu*SAS was not solved with metal bound. As previously mentioned in Section 5.1.4, the majority of active site residues are conserved between both species, however, there are some notable differences.

With the exception of ^{Nme}Asn184, each of the residues which form contacts to PEP in the 1XUZ structure are conserved and similarly positioned in the *Ssu*SAS structure. As previously mentioned, a phosphate ion is observed in each *Ssu*SAS active site. Shown in Figure 5.22, the phosphate moiety from Chain A of the *Ssu*SAS structure overlays well with the phosphate group of the PEP molecule in 1XUZ, which likely accounts for the similar positioning of the surrounding residues. No electron density was observed for a metal ion in the metal binding site of the *Ssu*SAS structure in either chain.

Depicted in Figure 5.22, ^{Nme}Asn184 forms a polar contact to the phosphate of PEP. In the *Ssu*SAS structure, ^{Nme}Asn174 is replaced by ^{Ssu}Thr171, which is not in a suitable position to interact with PEP. The shorter side-chain of threonine compared to asparagine may indicate that the contact observed in the *Nme*NANAS structure is not possible in *Ssu*SAS. This difference may account for the somewhat decreased binding affinity for PEP in *Ssu*SAS compared to *Nme*NANAS noted in Section 5.4.1. In addition, ^{Nme}Asn184 is fully conserved in all members of the NANAS clade, whereas ^{Ssu}Thr171 is a conserved threonine or serine in the LegS clade.

As mentioned in Section 5.1.4, each of the sugar binding residues in *Nme*NANAS are conserved in *Ssu*SAS with the exception of ^{Nme}Asn74 and ^{Nme}Phe288. ^{Nme}Asn74 is contributed by the β2-α2 loop in *Nme*NANAS which, as discussed in Chapter 4, is a region which varies in sequence between each clade of the SAS family. Whilst ^{Nme}Asn74 is not conserved in *Ssu*SAS, a bulky tyrosine (^{Ssu}Tyr54) protrudes into the active site from a similar position (Figure 5.22). ^{Ssu}Tyr54 is conserved within the LegS clade whilst ^{Nme}Asn74 is only conserved within the NANAS clade. Similarly, ^{Nme}Phe288 which is conserved in the NANAS clade, is replaced by ^{Ssu}Val275, which is conserved in the LegS clade (Chapter 4).

Furthermore, the conserved ManNAc binding residues are slightly adjusted in the *Ssu*SAS structure when compared with their counterparts in 1XUZ. Depicted in Figure 5.22, ^{Ssu}Glu37, ^{Ssu}Asp234 and ^{Ssu}Arg301 have shifted by 0.8 Å, 2.2 Å and 1.3 Å respectively from the corresponding residues in the *Nme*NANAS structure. The differences in positioning of the aforementioned residues are fairly minor and a similar degree of movement is observed in un-liganded dynamics of the *Nme*NANAS structure (Chapter 3). The alternative posturing of the *Ssu*SAS ManNAc binding residues may be caused by the absence of ManNAc and PEP in the active site. Furthermore, the position of ^{Ssu}Arg301 is likely influenced by the movement of the AFPL domain which, as depicted in Figure 5.17A, has moved closer to the catalytic domains in the *Ssu*SAS structure when compared to 1XUZ. Alternatively, the position of ^{Ssu}Arg301 may be influenced by the presence of the bound glycerol molecule depicted in Figure 5.21.

5.6.3 Small-angle X-ray scattering

SAXS was used to assess the solution structure of *Ssu*SAS. *Ssu*SAS was analysed in the presence of 1 mM Mn^{2+} and 1 mM PEP (Figure 5.23). The SAXS profile of *Ssu*SAS was determined using the methods outlined in Sections 8.5.8 and 8.5.9. Displayed in Figure 5.23A, the scattering profile of *Ssu*SAS in the presence of Mn^{2+} and PEP is compared to the scattering profile of *Nme*NANAS under the same conditions. The experimental scattering obtained for *Ssu*SAS is also compared with CRYSOL (*153*) fits generated from the ligand bound *Nme*NANAS crystal structure 1XUZ (*112*) and the ligand-free crystal structure of *Ssu*SAS outlined in Section 5.6.2. The structural parameters calculated for *Ssu*SAS via SAXS, are compiled in Table 5.10 and compared against values determined for liganded *Nme*NANAS.

	***Nme*NANAS**	***Ssu*SAS**
I_0 (cm^{-1})	0.082 ± 0.001	0.075 ± 0.001
R_g (Å)	33.0 ± 0.5	31.9 ± 0.3
D_{max} (Å)	116	113
Porod volume ($Å^3$)	121,724	115,631
Estimated mass (kDa)	85	81
Theoretical mass of dimer (kDa)	77	75

Table 5.10 SAXS structural parameters determined for *Ssu*SAS in the presence of 1 mM Mn^{2+} and 1 mM PEP, compared with structural parameters for *Nme*NANAS determined under the same conditions. SAXS data obtained at a wavelength of 1.0332 Å.

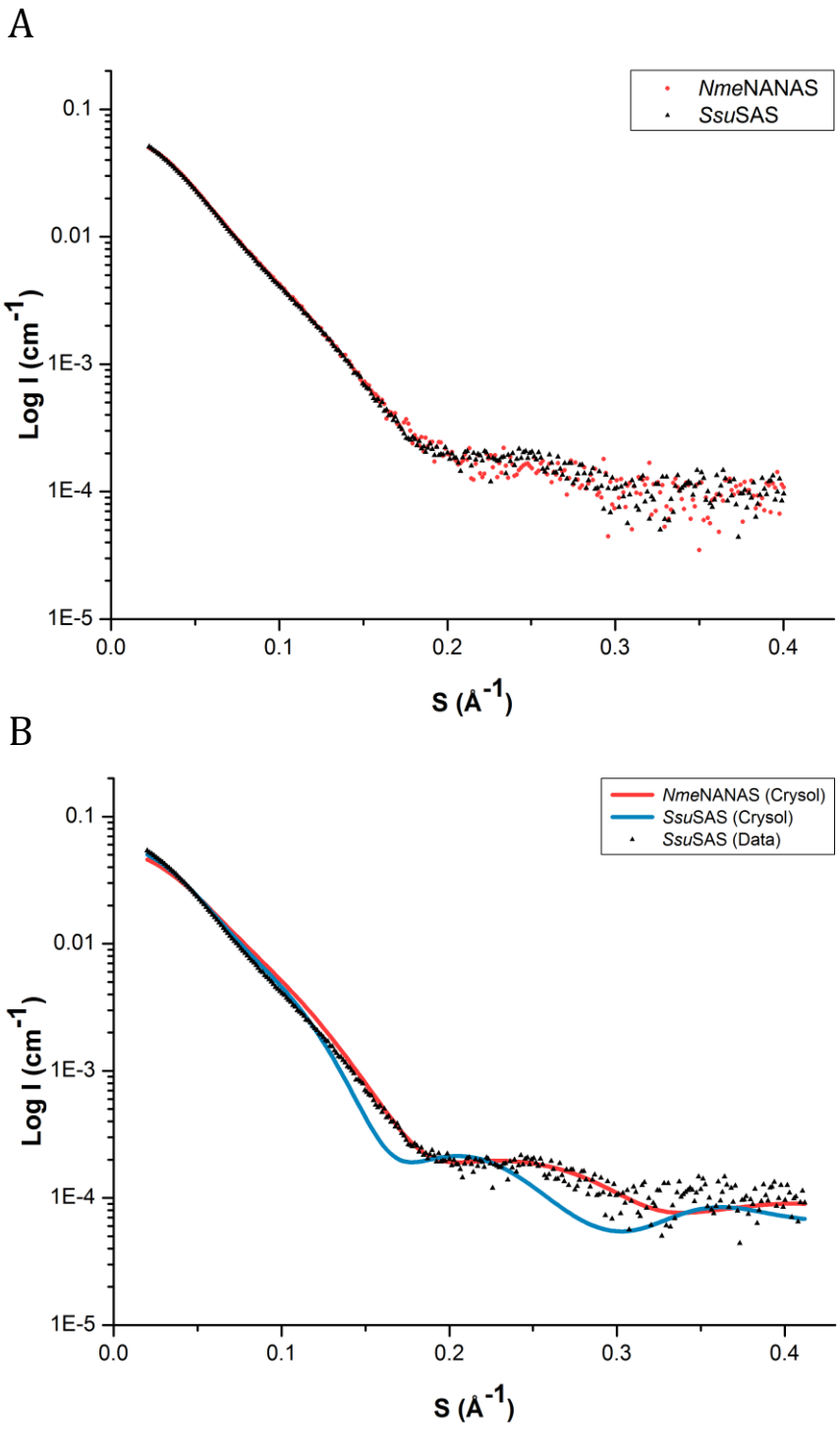


Figure 5.23 SAXS profile of *Ssu*SAS in the presence of 1 mM Mn^{2+} and 1 mM PEP. (A) Comparison of SAXS data for *Ssu*SAS (black triangles) with *Nme*NANAS wild-type (red circles). (B) CRYSOL fits of liganded *Ssu*SAS SAXS data with theoretical scattering of *Nme*NANAS wild-type generated from crystallographic coordinates 1XUZ (red line) and theoretical scattering generated from the *Ssu*SAS crystal structure detailed in Section 5.6 (blue line). SAXS data was collected using approximately 10 mg mL^{-1} protein in 30 mM triethanolamine-HCl buffer at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

The structure parameters obtained for *Ssu*SAS via SAXS are similar to those determined for liganded *Nme*NANAS (Table 5.10). Depicted in Figure 5.23A, the scattering profile of *Ssu*SAS is also similar to that observed for liganded *Nme*NANAS, indicating that the two enzymes have similar structures in solution. The molecular mass of *Ssu*SAS was calculated from SAXS data using SAXS MoW (*154*), and was estimated as 81 kDa, which is similar to that expected for dimeric enzyme (Table 5.10). The quaternary state of *Ssu*SAS determined from SAXS correlates with the results obtained from analytical SEC of *Ssu*SAS, which also indicates that *Ssu*SAS exists as a dimer in solution (Section 5.5.3).

Displayed in Figure 5.23B, the SAXS data obtained for *Ssu*SAS fits moderately well to the CRYSOL profile generated from the liganded *Nme*NANAS crystal structure 1XUZ and has a Chi value of 2.05. However, the fit to the CRYSOL profile generated from the ligand-free *Ssu*SAS crystal structure is significantly lower and has a Chi value of 6.06. The theoretical scattering profile generated from the *Ssu*SAS crystal structure is quite different to the profile for the *Nme*NANAS structure and is likely caused by the differences in conformation between the two crystal structures (as can be observed in Figure 5.17A). The conformational differences between the *Nme*NANAS and *Ssu*SAS structures are discussed in detail in Section 5.6.2.

Described in Chapter 3, wild-type *Nme*NANAS undergoes a conformational change in the presence of Mn^{2+} and PEP that is detectable by SAXS, and is linked to a H-bonding network that involves the rearrangement of the ManNAc binding site upon ligation of PEP. The *Ssu*SAS crystal structure does not have ligands bound and is possibly in a conformation which does not accurately reflect the state of the enzyme when Mn^{2+} and PEP are present. Given that the experimental scattering obtained for *Ssu*SAS was obtained in the presence of Mn^{2+} and PEP, it is unsurprising that the data fits better to the theoretical scattering of ligand bound *Nme*NANAS than to that of the ligand free *Ssu*SAS CRYSOL.

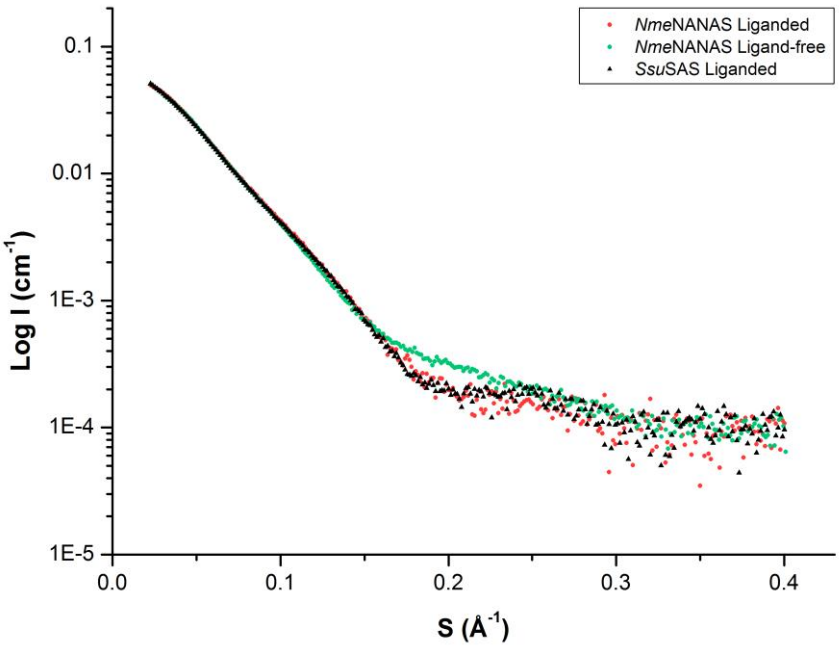


Figure 5.24 SAXS profile of *Ssu*SAS in the presence of 1 mM Mn^{2+} and 1 mM PEP compared with liganded and ligand-free scattering for *Nme*NANAS wild-type. SAXS data for liganded *Ssu*SAS is in black, liganded *Nme*NANAS wild-type in red and ligand-free *Nme*NANAS wild-type in green.

Displayed in Figure 5.24, the experimental scattering determined for *Ssu*SAS in the presence of Mn^{2+} and PEP is analogous to that obtained for liganded *Nme*NANAS and not ligand-free *Nme*NANAS. This likely indicates that a conformational change occurs upon binding of PEP to *Ssu*SAS, similar to what is observed for *Nme*NANAS (Chapter 3). The conformational change observed in the presence of Mn^{2+} and PEP would also explain the increased thermal stability observed for *Ssu*SAS with Mn^{2+} and PEP described in Section 5.4.3.

The results obtained from SAXS and DSF as well as the PEP ITC profile for *Ssu*SAS, correlate with the observations for *Nme*NANAS outlined in Chapter 3 and indicate that the mechanism for pre-organisation of the *Nme*NANAS ManNAc binding site is retained in *Ssu*SAS. Described in Chapter 3, residues ^{Nme}Glu134 and ^{Nme}Glu282 were identified to play a key role in the hydrogen bonding network involved with ManNAc pre-organisation. Depicted in the sequence alignment between *Nme*NANAS and *Ssu*SAS in Figure 5.3, both ^{Nme}Glu134 and ^{Nme}Glu282 are conserved in *Ssu*SAS. The conservation of these residues is further confirmation that the ManNAc pre-organisation mechanism is retained in *Ssu*SAS.

5.6.4 Analytical ultracentrifugation

Both analytical SEC (Section 5.5.3) and SAXS (Section 5.6.3) indicated that *Ssu*SAS exists as a dimeric protein. However, both of these techniques were run at high protein concentrations (2 - 10 mg/mL) and may not accurately portray the quaternary state at lower concentrations of *Ssu*SAS. Analytical ultracentrifugation (AUC) allows monitoring of quaternary structure at much lower protein concentrations than either SAXS or analytical SEC. AUC was thus used to determine if the quaternary state of *Ssu*SAS was concentration dependent by running samples at 0.1 mg/mL, 0.5 mg/mL and 1.0 mg/mL. Displayed in Figure 5.25, data was fitted to a continuous size-distribution (c_s) model using the program SEDFIT (*150, 151*). Weight average molecular masses were calculated using weight average sedimentation coefficients and f/f_0 values from c_s fits. The method used for anlaysis is outlined in Section 8.5.7.

As displayed in Figure 5.25D, *Ssu*SAS eluted as a single species at each of the concentrations tested via AUC. The sedimentation coefficient peak corresponds to a molecular weight of 78 kDa which is very similar to that calculated for dimeric *Ssu*SAS (75 kDa). The results from AUC indicate that *Ssu*SAS is a stable dimer even at low concentrations and agrees with the results ascertained using analytical SEC and SAXS. Similar stability of dimeric protein is observed with wild-type *Nme*NANAS (Chapter 3).

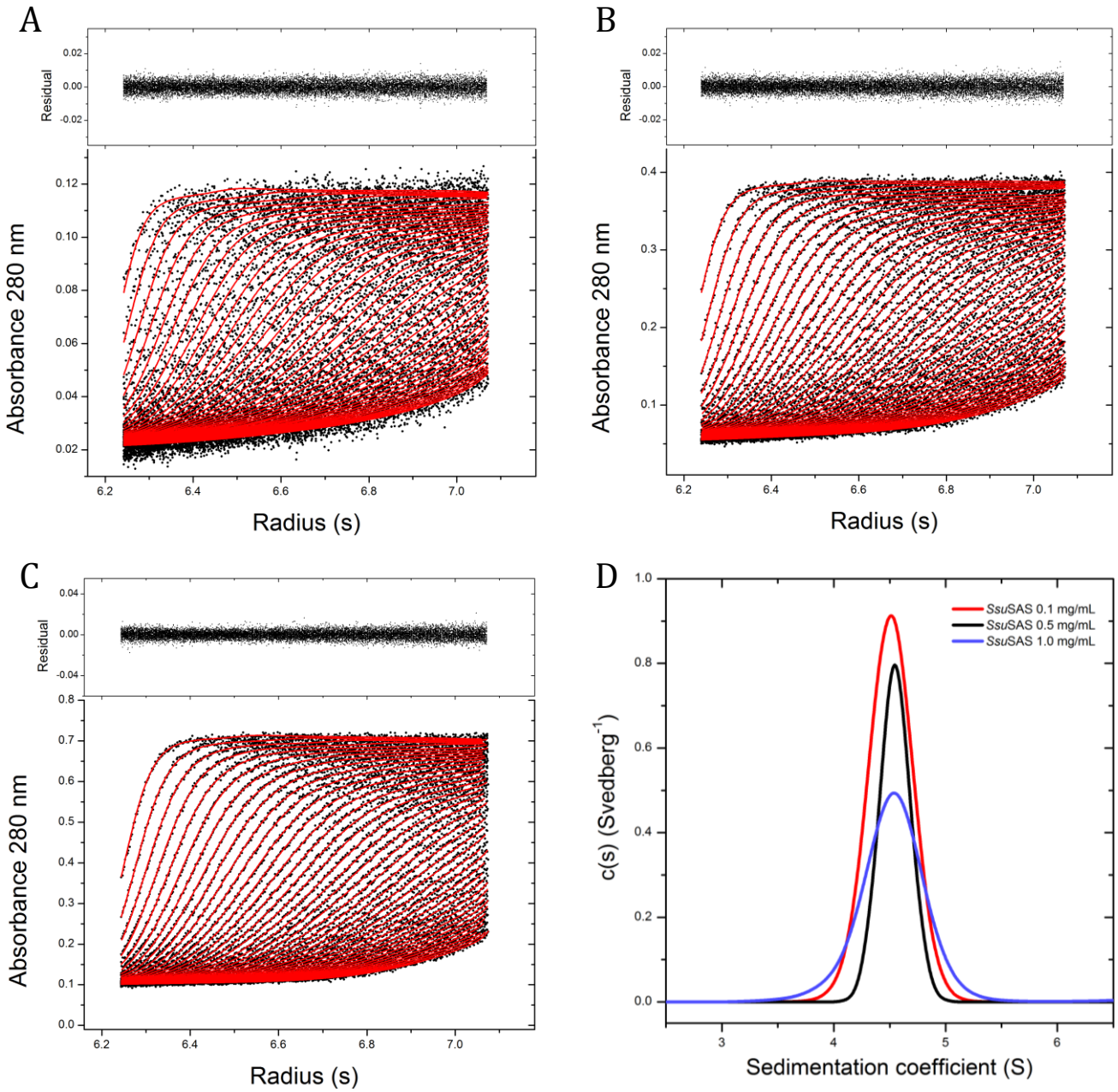


Figure 5.25 Analytical ultra centrifuge fits measured at 280 nm absorbance. (A) *Ssu*SAS at 0.1 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (B) *Ssu*SAS at 0.5 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (C) *Ssu*SAS at 1.0 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (D) Normalised c(s) profiles for *Ssu*SAS at 0.1 mg/mL (red), 0.5 mg/mL (black) and 1.0 mg/mL (purple).

5.6.5 Substrate docking

Since a structure of *Ssu*SAS could not be elucidated with substrates bound, docking of Mn^{2+}, PEP and ManNAc was attempted by Dr Wanting Jiao in order to determine the interactions within the *Ssu*SAS active site. Displayed in Figure 5.26 is the best pose afforded from docking experiments with ManNAc.

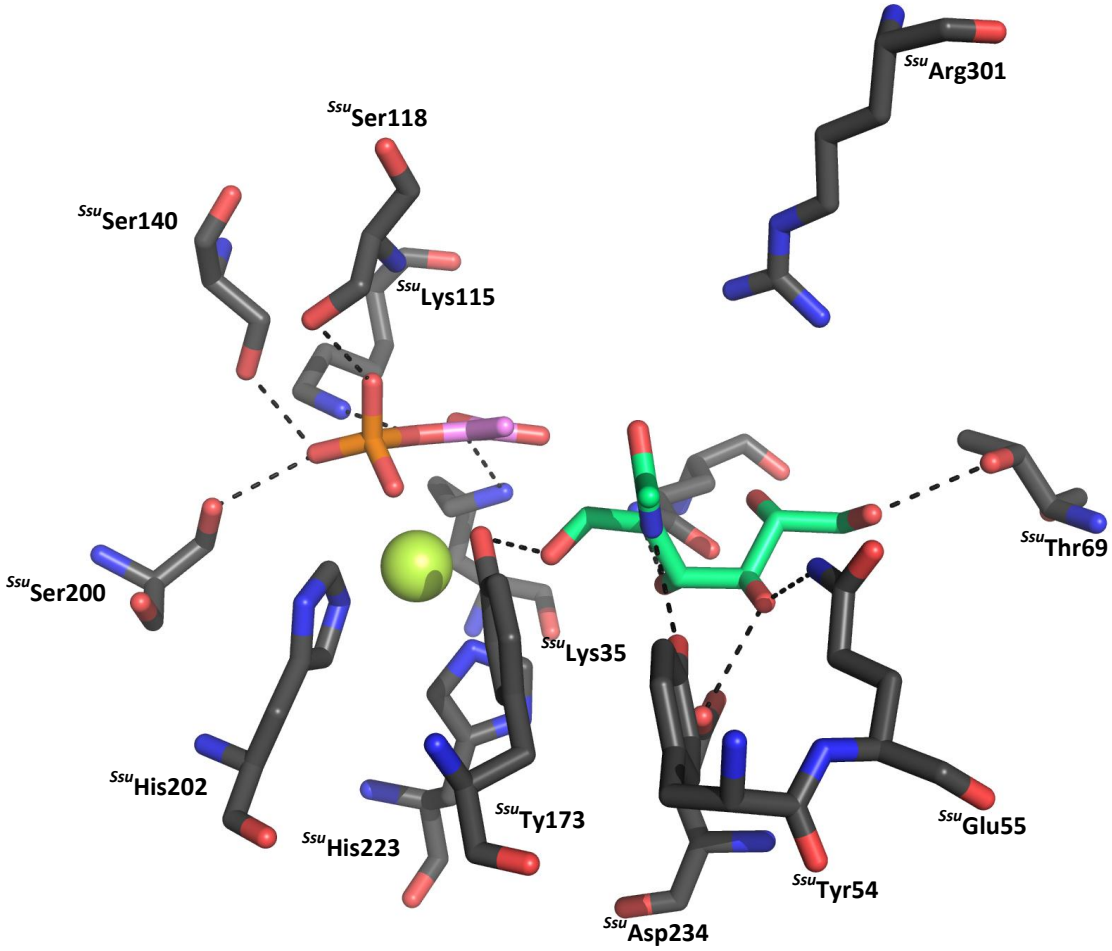


Figure 5.26 Best pose attained from docking of Mn^{2+}, PEP and ManNAc into the active site of *Ssu*SAS. *Ssu*SAS residues are shown in grey, with Mn^{2+} in yellow, PEP in pink and ManNAc in green. Polar contacts are depicted with black dashes.

Displayed in Figure 5.26, the docked PEP molecule forms similar contacts in *Ssu*SAS to *Nme*NANAS 1XUZ. Interestingly, the best pose obtained from docking of ManNAc was not a suitable reactive orientation. The docked ManNAc bound deep into the active site pocket, forming contacts with residues contributed from the β2-α2 loop of *Ssu*SAS such as ^{Ssu}Tyr54, ^{Ssu}Glu55 and ^{Ssu}Thr69. ^{Ssu}Tyr54, ^{Ssu}Glu55 are each fully conserved within the LegS clade, however, ^{Ssu}Thr69 is not (Figure 5.3).

The most significant difference in the positioning of the docked ManNAc in *Ssu*SAS compared to rManNAc in *Nme*NANAS 1XUZ, is the loss of the interaction between ^{Ssu}Arg301 and the *N*-acetyl group of ManNAc. As outlined in Chapter 2, *Nme*Arg314 plays an important role in the positioning of ManNAc for reaction with PEP. Displayed in Figure 5.26, ^{Ssu}Arg301 is 4.5 Å away from the *N*-acetyl group of ManNAc and is thus not within hydrogen bonding distance. Although the docking analysis of ManNAc did not provide a reactive pose for the sugar substrate, *Ssu*SAS was successfully characterised with NANAS activity (Section 5.3.1).

The unexpected binding pose for ManNAc is likely caused by the template structure used for docking. As discussed in Section 5.6.2 the *Ssu*SAS structure is in a different conformation to *Nme*NANAS which is probably due to the absence of PEP. In particular, the *Ssu*SAS AFPL domains are re-orientated and consequently the position of ^{Ssu}Arg301 is slightly adjusted.

As discussed in Section 5.1.4, *Ssu*SAS bears remarkable sequence similarity to members of the LegS clade of SAS enzymes. Unfortunately, the sugar substrate for Leg biosynthesis (2,4-diacetamido-2,4,6-trideoxymannose) is not commercially available and its synthesis is not trivial. Due to the unavailability of 2,4-diacetamido-2,4,6-trideoxymannose, *Ssu*SAS could not be tested for LegS activity. However, to determine if *Ssu*SAS could possibly bind 2,4-diacetamido-2,4,6-trideoxymannose, the LegS substrate was docked with Mn^{2+} and PEP into the active site of the *Ssu*SAS crystal structure (Figure 5.27).

Contrary to ManNAc, 2,4-diacetamido-2,4,6-trideoxymannose docks in a suitable reactive position in the active site of *Ssu*SAS (Figure 5.27). Both of the *N*-acetyl groups from the LegS substrate form interactions with *Ssu*Arg301 which, as aforementioned, is likely involved in steering the sugar substrate into a reactive stance. Polar contacts are also formed between ^{Ssu}Glu37, ^{Ssu}Tyr173 and ^{Ssu}Asp234 which are each conserved ManNAc binding residues identified in *Nme*NANAS 1XUZ. Unlike the docking of ManNAc, no interactions are observed between 2,4-diacetamido-2,4,6-trideoxymannose and the β2-α2 loop of *Ssu*SAS.

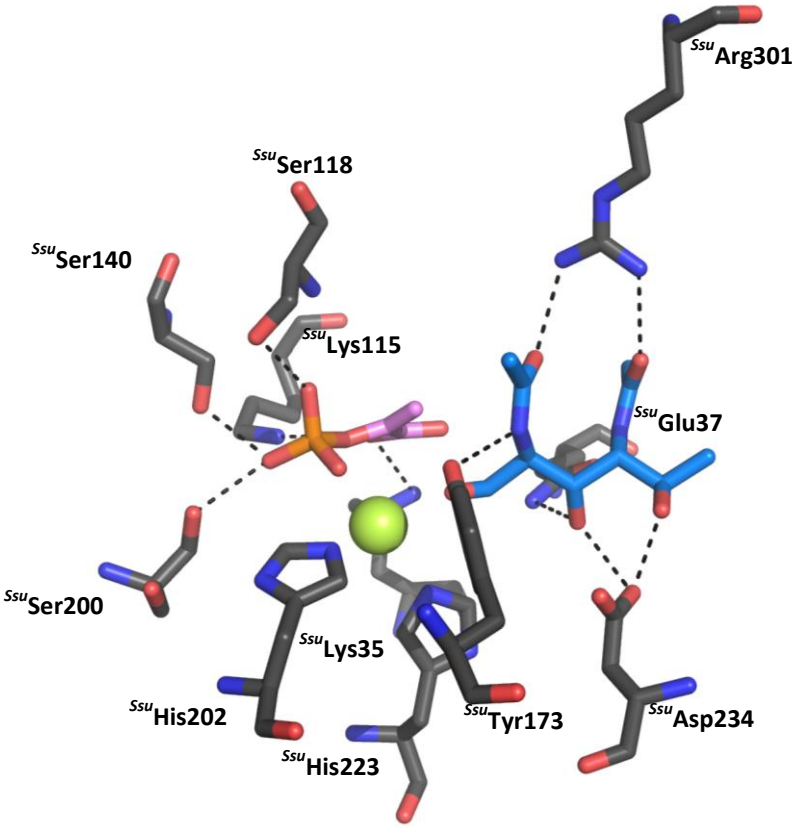


Figure 5.27 Best pose attained from docking of Mn^{2+}, PEP and 2,4-diacetamido-2,4,6-trideoxymannose into the active site of *Ssu*SAS. *Ssu*SAS residues are shown in grey, with Mn^{2+} in yellow, PEP in pink and 2,4-diacetamido-2,4,6-trideoxymannose in blue. Polar contacts are depicted with black dashes.

The results from docking suggest that *Ssu*SAS is capable of binding 2,4-diacetamido-2,4,6-trideoxymannose in a more favourable reactive orientation than ManNAc. Consequently, the docking results strongly suggest that *Ssu*SAS is capable of LegS activity. Additionally, the results from docking of 2,4-diacetamido-2,4,6-trideoxymannose in *Ssu*SAS produce almost identical interactions to those obtained for docking of the same compound with *Cje*LegS in Section 4.6.7. This correlates with the genealogical analysis outlined in Sections 5.1.3 and 5.1.4, indicating that *Ssu*SAS is a LegS. However, it should be noted that the results from substrate docking may be influenced by the starting positions of the active site residues in the *Ssu*SAS structure. Thus further experimental evidence would be required to validate LegS activity, i.e. kinetically assaying *Ssu*SAS with 2,4-diacetamido-2,4,6-trideoxymannose.

5.7 Summary of results

The SAS from *S. suis* was successfully expressed and purified by metal affinity and size exclusion chromatography. NANAS activity was confirmed by assaying *Ssu*SAS with PEP and ManNAc. *Ssu*SAS was revealed to have comparable kinetic parameters to those previously determined for wild-type *Nme*NANAS in Chapter 2. An optimal pH of 7 and optimal temperature of 40 °C was determined for NANAS activity.

The effects of metal ions upon the activity of *Ssu*SAS were analysed, with maximal activity observed for Co^{2+} followed by Mn^{2+}. Several metal ions were discovered to inhibit activity including Ca^{+}, Fe^{2+}, Zn^{2+} and Cd^{2+}. Activity of *Ssu*SAS could be abolished upon the addition of EDTA confirming metal dependence of catalysis. Although sugar substrate promiscuity was observed in the literature for the closely related enzyme *Sag*SAS (*105*), no activity could be ascertained for *Ssu*SAS with the same compounds.

The thermal stability of *Ssu*SAS was analysed in the presence of various metal ions. Similar to the trends observed with metal activation, Co^{2+} was found to be the most stabilising metal ion followed by Mn^{2+}. DSF of *Ssu*SAS with combinations of PEP, ManNAc and rManNAc revealed a similar thermal melting profile to that of wild-type *Nme*NANAS, where a significant thermal shift was observed upon the addition of PEP.

ITC profiles of *Ssu*SAS with Mn^{2+} and PEP were also similar to *Nme*NANAS wild-type. Furthermore, results obtained from SAXS indicate that *Ssu*SAS is capable of undergoing a similar conformational change to that observed for wild-type *Nme*NANAS in the presence of PEP and Mn^{2+}. Outlined in Chapter 3, PEP binding pre-organises the ManNAc binding site via a hydrogen bonding network which links the catalytic and AFPL domains. The H-bond network residues *Nme*Glu134 and *Nme*Glu282 are both conserved in *Ssu*SAS, indicating that a similar mechanism is possible.

The molecular mass of *Ssu*SAS was successfully ascertained using MS and the protein was determined to be folded in solution using CD. *Ssu*SAS was confirmed to be dimeric by SAXS, analytical SEC and AUC.

*Ssu*SAS was successfully crystallised and a structure with phosphate and glycerol bound was elucidated. *Ssu*SAS was determined to be a domain swapped homodimer with each monomer unit comprised of a $(\beta/\alpha)_8$ barrel catalytic domain with a short linker sequence to a C-terminal AFPL domain, similar to the previously refined structures of *Nme*NANAS (*112, 113*).

Slight structural differences were observed between *Ssu*SAS and *Nme*NANAS. The AFPL domains in the *Ssu*SAS structure were positioned closer to the catalytic domains when compared to the *Nme*NANAS structure. As a result, ^{Ssu}Arg301 which is contributed by the AFPL domain of the opposing monomer was repositioned in the active site of *Ssu*SAS when compared with the congruent residue in *Nme*NANAS 1XUZ.

Since a crystal structure of *Ssu*SAS could not be obtained with substrates bound, molecular docking was used to determine possible catalytic site interactions. Docking of ManNAc into the active site of *Ssu*SAS did not afford a reactive pose, however, docking of the LegS substrate 2,4-diacetamido-2,4,6-trideoxymannose did. The docking results obtained for *Ssu*SAS with 2,4-diacetamido-2,4,6-trideoxymannose imply that the substrate for Leg biosynthesis can be reacted with PEP. Furthermore, *Ssu*SAS shares high sequence identity to members of the LegS clade of SASs, indicating the enzyme is likely to have LegS activity.

5.8 Discussion

The initial aim of Chapter 5 was to uncover substrate promiscuity in the SAS family. The SAS from *S. suis* was selected for investigation due to its high sequence similarity to the SAS from *S. agalactiae,* which was previously observed to utilise a variety of ManNAc analogues (*105*). Whilst the SAS from *S. suis* was determined to have NANAS activity, no detectable activity was observed with any of the ManNAc analogues tested. The optimal pH, temperature and metal ion for activity of *Ssu*SAS were also identified. The DSF, ITC and SAXS profiles of *Ssu*SAS in the presence of PEP were similar to those observed for wild-type *Nme*NANAS. Furthermore, the residues involved in the hydrogen bonding network of *Nme*NANAS are conserved in *Ssu*SAS. Given the aforementioned observations, it is highly likely that the *Nme*NANAS mechanism of ManNAc binding-site pre-organisation is also conserved in *Ssu*SAS.

A crystal structure of *Ssu*SAS was elucidated to a resolution of 2.3 Å. *Ssu*SAS was determined to be a domain swapped homodimer with a similar overall architecture to that previously observed for wild-type *Nme*NANAS (*112, 113*). Although it was observed to have NANAS activity, *Ssu*SAS was positioned in the LegS clade of the phylogram described in Chapter 4. Inspection of the *Ssu*SAS sequence revealed that the region pertaining to the β2-α2 loop was indeed more closely related to members of the LegS clade than NANAS clade.

Molecular modelling was used to predict the possible interactions of both ManNAc and 2,4-diacetamido-2,4,6-trideoxymannose in the active site of the *Ssu*SAS crystal structure. The docking results obtained for *Ssu*SAS with 2,4-diacetamido-2,4,6-trideoxymannose strongly implied that that the substrate for Leg biosynthesis could be reacted with PEP. Furthermore, the binding poses attained for 2,4-diacetamido-2,4,6-trideoxymannose in the *Ssu*SAS crystal structure were comparable to those observed for ManNAc.

Chapter 6

Generation and characterisation of *Homo sapiens* and *Neisseria meningitidis* AFPL domain chimeras

6.1 Background

6.1.1 Mammalian and bacterial NANA biosynthesis

The sialic acid, NANA is commonly integrated as a terminal residue on mammalian cell surface glycoconjugates (*177*). NANA plays a significant role in both cellular recognition and cell adhesion processes in mammals (*2, 178*). NANA is also produced by neuroinvasive bacteria such as *Neisseria meningitidis*, the primary causative agent of bacterial meningitis (*179*) and by *C. jejuni*, the primary causative agent of food-borne gastroenteritis (*88*). These select pathogenic organisms incorporate NANA in their cell surface, mimicking the surface of a mammalian cell. This cellular mimicry enables the bacteria to evade the host's immune response (*180*).

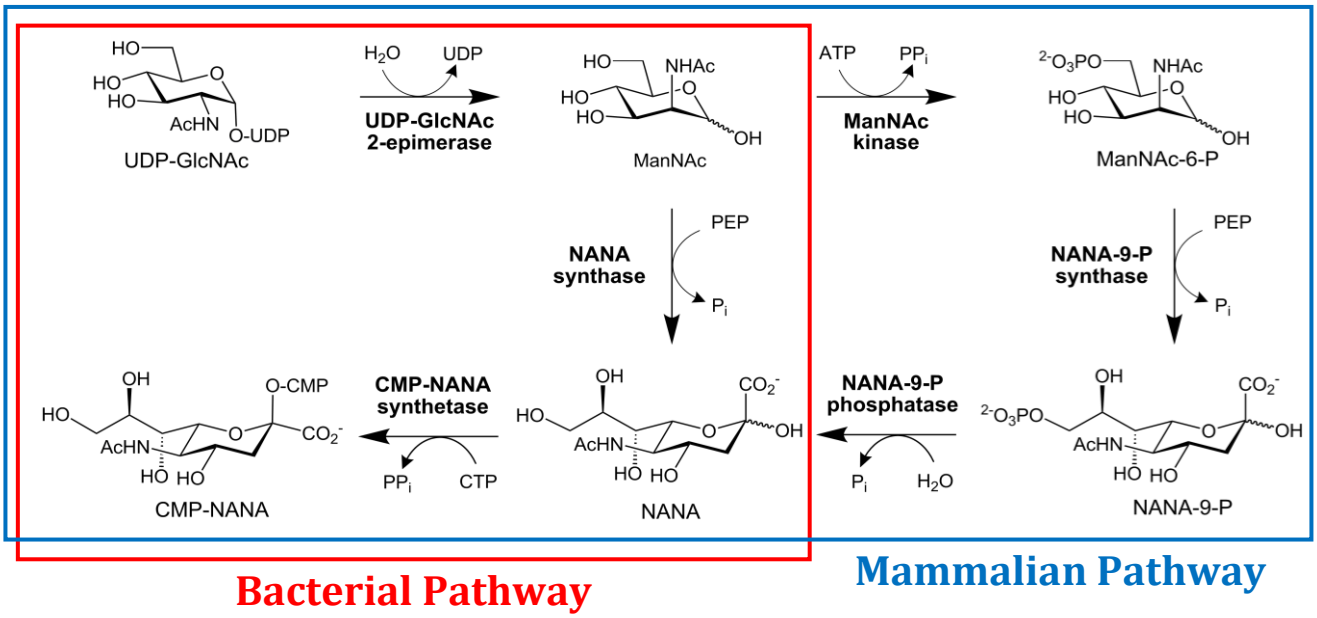


Figure 6.1 The biosynthetic pathways of CMP-NANA in bacteria and mammals. The reaction pathway in bacteria is boxed in red and the pathway in mammals is boxed in blue. In bacteria, ManNAc is directly converted to NANA by NANA synthase whereas NANA is produced via phosphorylated intermediates in the mammalian pathway.

The biosynthesis of NANA in bacteria and mammals differs slightly (Figure 6.1) (*5*), though both pathways involve the formation of ManNAc from UDP-GlcNAc. This reaction is catalysed by UDP-GlcNAc 2-epimerase, which hydrolyses the glycosidic-UDP linkage of UDP-GlcNAc and inverts the stereochemistry at C2. In the bacterial pathway, the subsequent reaction is the aldol-like condensation of PEP and ManNAc to form NANA which is catalysed by NANAS (*112*).

In the mammalian pathway, ManNAc is first phosphorylated at C-6 by ManNAc kinase to form ManNAc-6-P. ManNac-6-P is then reacted with PEP by NANA-9-PS to form NANA-9-P (*103*). Finally, NANA-9-P is dephosphorylated by NANA-9-P phosphatase to produce NANA. In both bacteria and mammals, NANA is activated by CMP-NANA synthetase which utilises cytosine triphosphate (CTP) and NANA to produce CMP-NANA. CMP-NANA is subsequently incorporated into cell surface glycoconjugates (*181*).

6.1.2 Goals of this research

As described in Chapters 2 and 3 of this thesis, the AFPL domain plays an important role in the binding and positioning of ManNAc. The AFPL domain is linked with the pre-organisation of the ManNAc binding site via a hydrogen bonding network and provides an essential arginine residue that helps steer ManNAc into a reactive position. The AFPL domain sequence in *Hsa*NANA-9-PS is significantly different to that from *Nme*NANAS and does not contain the aforementioned critical arginine.

The aim of the work done in this chapter was to determine whether the function of the *Hsa*NANA-9-PS and *Nme*NANAS AFPL domains were similar and could be interchanged. This was to be done by generating and characterising two chimeric proteins. Shown in Figure 6.2, the first chimera was designed to comprise of the *N. meningitidis* catalytic domain joined to the *H. sapiens* linker and AFPL domain (*Nme*/*Hsa* chimera) and the second chimera was designed to comprise of the *H. sapiens* catalytic domain joined to the *N. meningitidis* linker and AFPL domain (*Hsa*/*Nme* chimera).

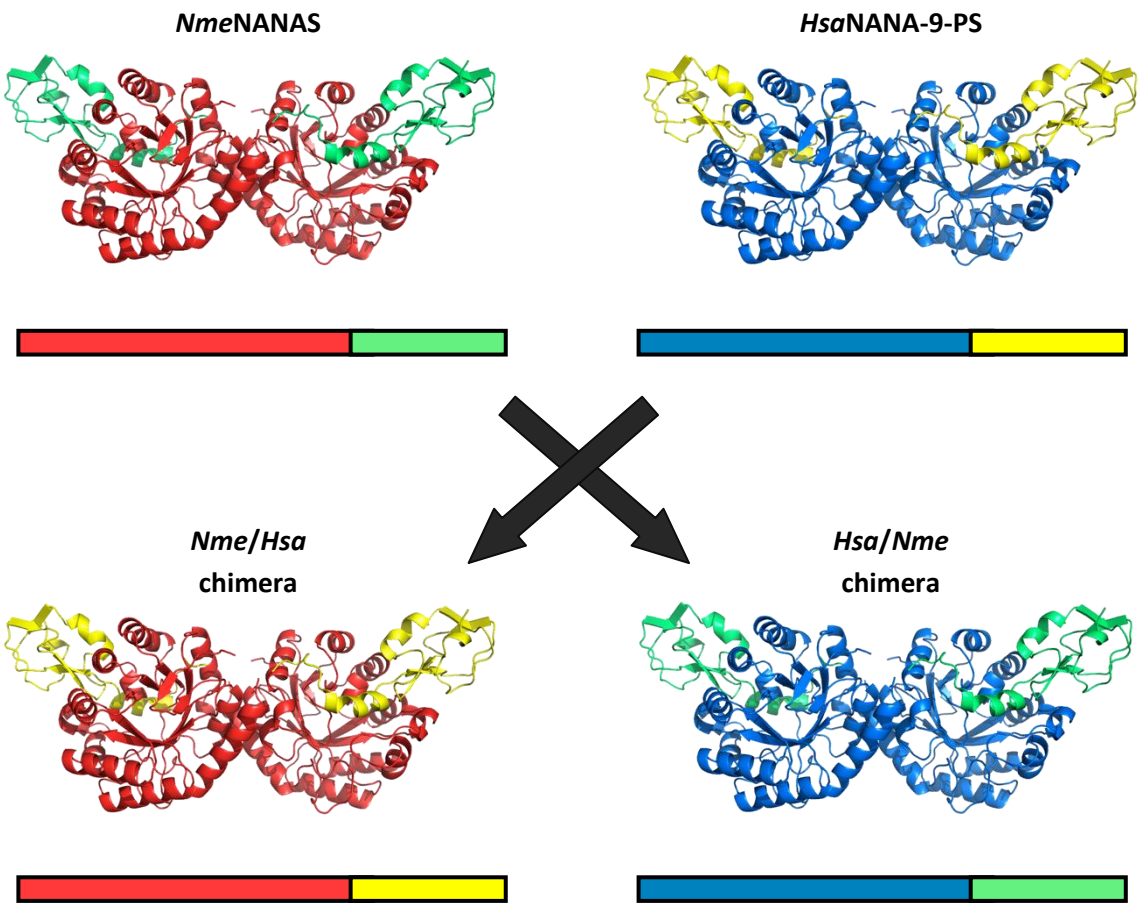


Figure 6.2 Schematic depicting NANA/NANA-9-PS chimera constructs. The *Nme*NANAS catalytic domain is depicted in red, *Nme*NANAS AFPL- domain in green, *Hsa*NANA-9-PS catalytic domain in blue and *Hsa*NANA-9-PS AFPL domain in green. The *Nme*/*Hsa* chimera is comprised of the *Nme*NANAS catalytic domain and *Hsa*NANA-9-PS linker and AFPL domain. The *Hsa*/*Nme* chimera is comprised of the *Hsa*NANA-9-PS catalytic domain and *Nme*NANAS linker and AFPL domain.

6.2 Sequence and Structural Comparison of *Nme*NANAS and *Hsa*NANA-9-PS

6.2.1 Sequence alignment of *Nme*NANAS and *Hsa*NANA-9-PS

In order to determine the relationship between *Nme*NANAS and *Hsa*NANA-9-PS, a sequence alignment was generated using ClustalW2 (*155, 156*). The sequence identity between *Nme*NANAS and *Hsa*NANA-9-PS was determined to be 27.9 %. The catalytic domains of *Nme*NANAS and *Hsa*NANA-9-PS share 29.8 % identity, whilst the AFPL domains have a lower identity of 25.9 %. As shown in Figure 6.3, the β2-α2 loop of *Hsa*NANA-9-PS is slightly longer than that of *Nme*NANAS. Furthermore, *Hsa*NANA-9-PS contains an additional eight C-terminal residues when compared to the sequence for *Nme*NANAS.

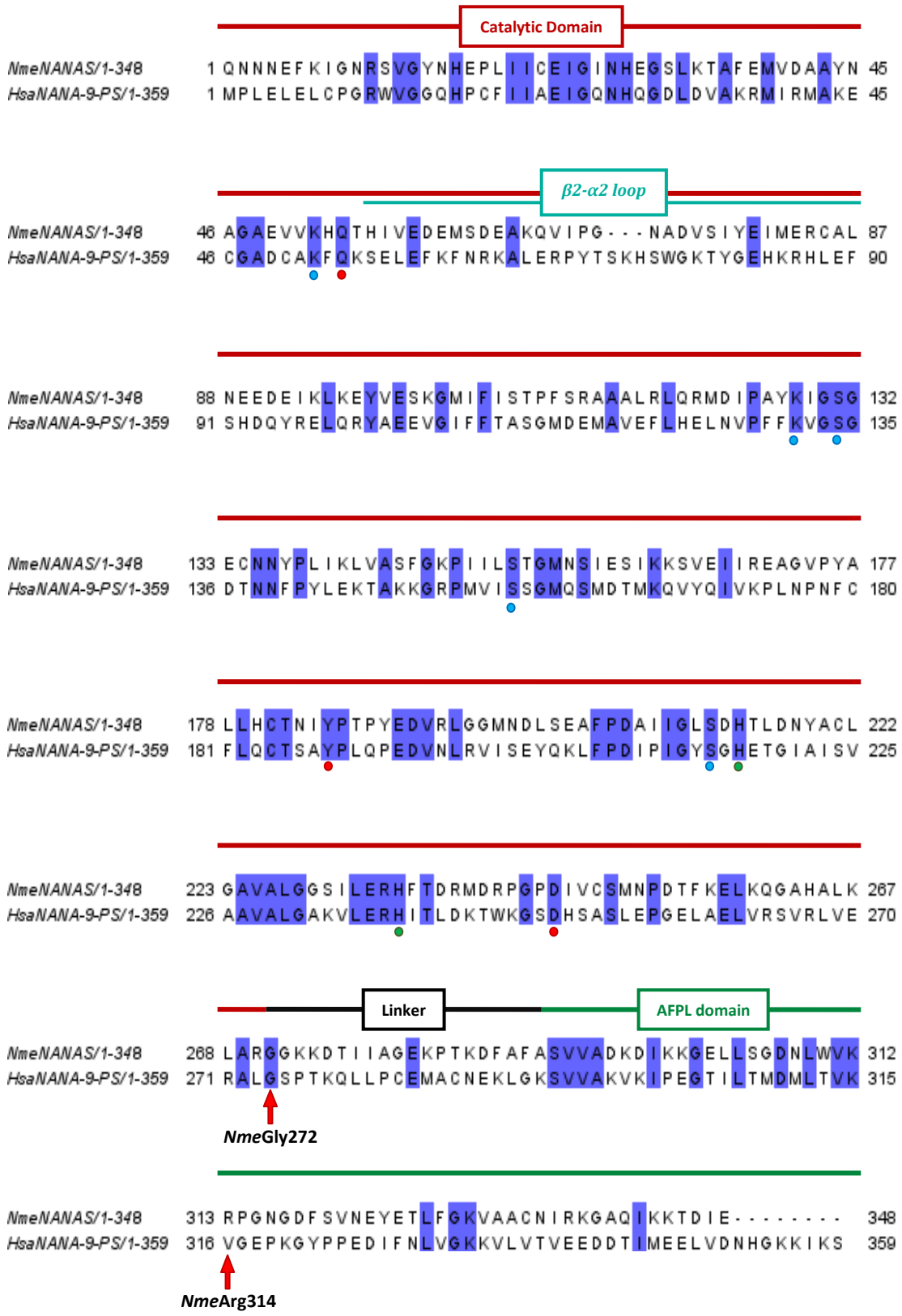


Figure 6.3 Sequence alignment of *Nme*NANAS and *Hsa*NANA-9-PS displayed using Jalview (*159*). Residues highlighted in blue indicate conservation across the SAS family. Conserved active site residues from *Nme*NANAS are marked with dots; metal binding residues in green, PEP binding residues in blue and ManNAc binding residues in red. Lines indicate sequence pertaining to *Nme*NANAS domains; catalytic barrel in red, β2-α2 loop in blue, linker in black and AFPL domain in green. The residue and domain annotation is based on the *Nme*NANAS crystal structure (PDB code (1XUZ) (*112*)).

As depicted in Figure 6.3, several of the key catalytic site residues are conserved between both *Nme*NANAS and *Hsa*NANA-9-PS. The *Nme*NANAS metal binding residues His215 and His236 are both conserved in *Hsa*NANA-9-PS. The *Nme*NANAS PEP binding residues Lys53, Lys129, Ser132, Ser154 and Ser213 are each conserved in *Hsa*NANA-9-PS. Thr110 is substituted by a serine residue in *Hsa*NANA-9-PS, however as detailed in Section 4.2.2, this position is conserved as either threonine or serine in all bacterial SASs. Asn184 is not conserved in *Hsa*NANA-9-PS either, but this residue is only conserved in the bacterial NANAS clade (Chapter 4).

The ManNAc binding residues Gln55, Tyr186 and Asp247 are each conserved in *Hsa*NANA-9-PS, however Asn74, Phe288 and Arg314 are not. The observation that Asn74 is not conserved in *Hsa*NANA-9-PS is unsurprising given that Asn74 is contributed by the β2-α2 loop. As discussed in Section 4.2.2, the β2-α2 loop varies significantly between different SAS clades and is likely involved in sugar substrate recognition. Shown in Figure 6.4, Asn74 interacts with the C6 hydroxyl of ManNAc in *Nme*NANAS. As aforementioned, *Hsa*NANA-9-PS utilises ManNAc-6-P and not ManNAc. The β2-α2 loop of *Hsa*NANA-9-PS would therefore be proximal to the C6 phosphate of ManNAc-6-P. Shown in Figure 6.5, the β2-α2 loop of *Hsa*NANA-9-PS contains far more positive residues than the corresponding sequence in *Nme*NANAS. This correlates with the electrostatic requirements for interacting with the negatively charged phosphate moiety.

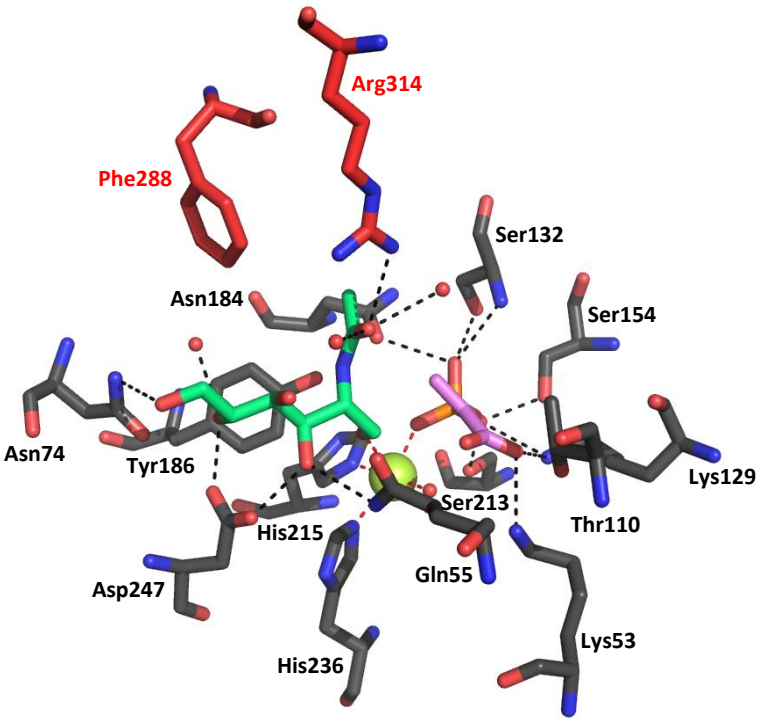


Figure 6.4 Close up of *Nme*NANAS active site. Residues contributed by the catalytic domain are shown in grey and residues from linker and AFPL domain in red. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are shown in red. H-bonds are depicted with black dashes and dipolar bonds with red dashes (PDB code (1XUZ) (*112*)).

Figure 6.5 Sequence alignment of *Nme*NANAS and *Hsa*NANA-9-PS β2-α2 loop displayed using Jalview (*159*). Residues highlighted in red indicate positively charged side-chains at neutral pH.

It is also unsurprising that Phe288 is not conserved in *Hsa*NANA-9-PS, as it is only conserved in members of the bacterial NANAS family and does not directly interact with ManNAc in the *Nme*NANAS structure (Figure 6.4). Surprisingly, the AFPL domain residue Arg314 is not conserved in *Hsa*NANA-9-PS. Arg314 interacts with the *N*-acetyl group of ManNAc and has been previously identified as an essential component for catalysis in *Nme*NANAS (*114*). As outlined in Chapter 2, Arg314 steers ManNAc into a reactive position with PEP. As indicated in Figure 6.3, Arg314 is replaced by a valine in *Hsa*NANA-9-PS, which is fully conserved in mammalian SASs. It should be noted that less than 30 mammalian SAS sequences are currently available for comparison.

6.2.2 Structure of *H. sapiens* NANA-9-PS

Currently there is no complete crystal structure of *Hsa*NANA-9-PS available however the structure of the *Hsa*NANA-9-PS AFPL domain has been elucidated by NMR (*115*). Displayed in Figure 6.6, the AFPL domain of *Hsa*NANA-9-PS shows high structural similarity to the AFPL domain of *Nme*NANAS (RMSD = 0.40 Å).

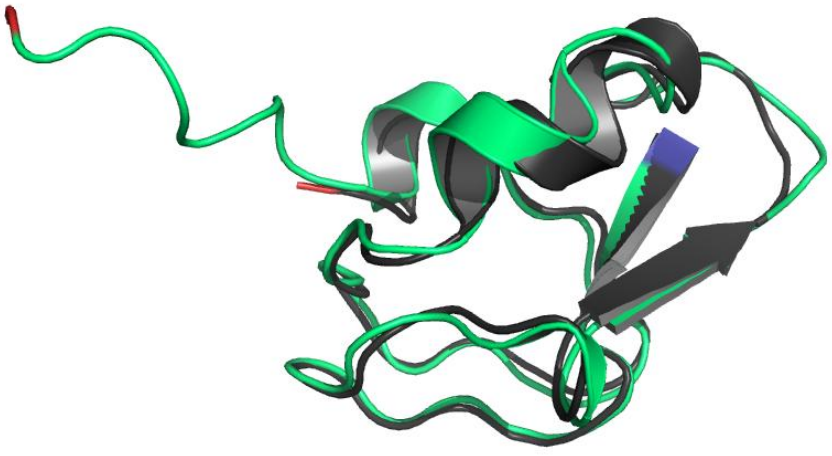


Figure 6.6 Alignment of *Nme*NANAS AFPL domain (grey) with *Hsa*NANA-9-PS AFPL domain (green). N-terminal residues are coloured blue and C-terminal residues are coloured red. *Nme*NANAS AFPL domain structure generated from crystallographic coordinates 1XUZ (*112*) and *Hsa*NANA-9-PS AFPL domain from coordinates 1WVO (*115*).

To determine a more complete representation of the *Hsa*NANA-9-PS structure, a molecular model was generated by Dr Wanting Jiao. The ligand bound *Nme*NANAS crystal structure (PDB code 1XUZ) was used as a template for the *Hsa*NANA-9-PS model (Figure 6.7). ManNAc-6-P was docked into the active site of the *Hsa*NANA-9-PS model to ascertain the key interactions between the enzyme and substrates. The best ligand pose determined through docking is displayed in Figure 6.7B.

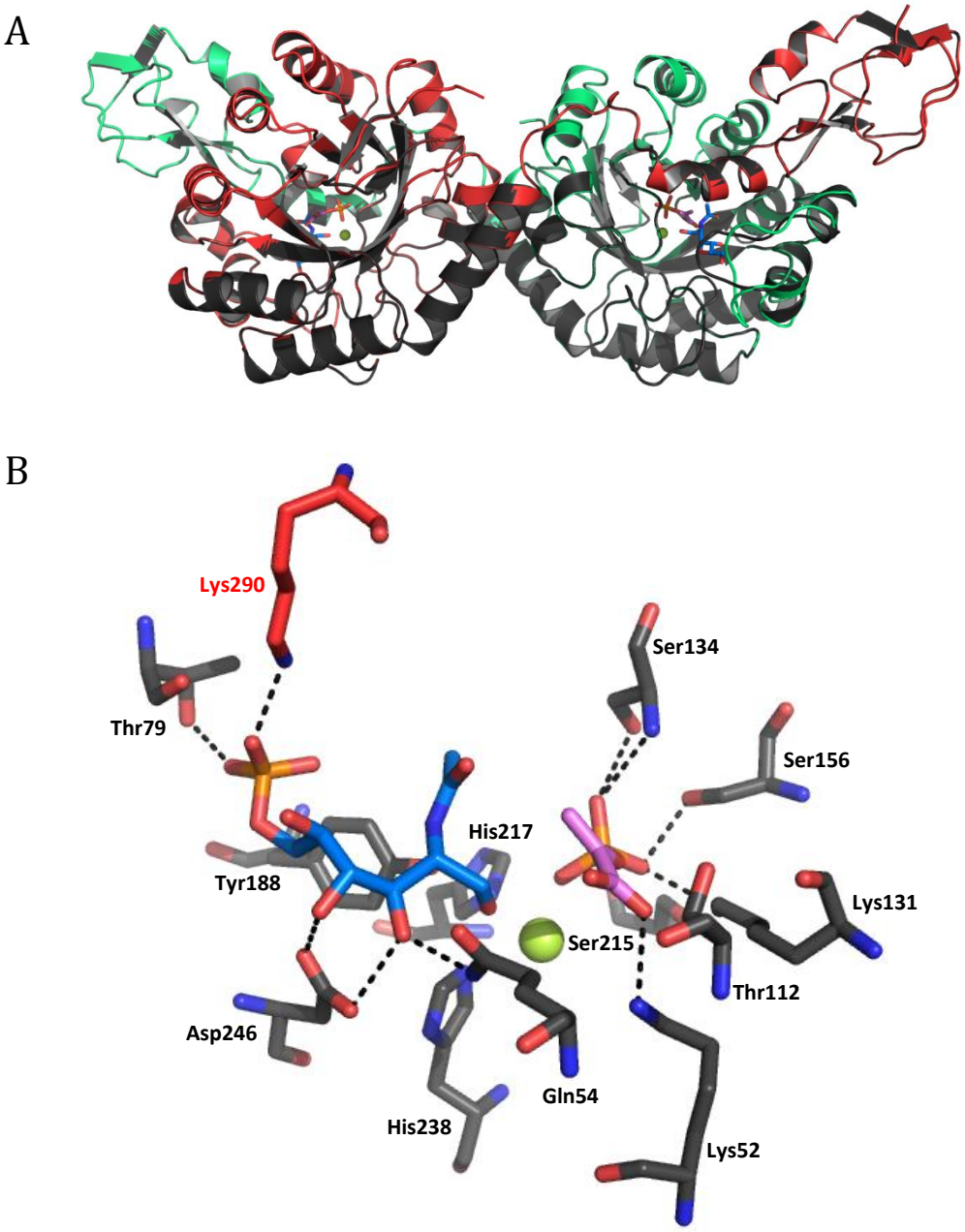


Figure 6.7 (A) *Hsa*NANA-9-PS dimer aligned to *Nme*NANAS dimer. Both chains of *Nme*NANAS are shown in black, chain A of *Hsa*NANA-9-PS is in red and chain B is in green. (B) Close up of *Hsa*NANA-9-PS active site with substrates docked. Residues contributed by the catalytic domain are shown in grey and residues from linker and AFPL domain in red. ManNAc-6-P is coloured blue, PEP pink, Mn^{2+} yellow. H-bonds are depicted with black dashes (PDB code (1XUZ) (*112*)).

As expected, the model of the *Hsa*NANA-9-PS dimer aligns well with the crystal structure of *Nme*NANAS and has an RMSD = 0.1 Å (Figure 6.3A). The overall fold of each *Hsa*NANA-9-PS monomer is maintained, with the catalytic barrel, linker and AFPL domains each structurally similar to the template fold.

The docking of ManNAc-6-P, PEP and Mn^{2+} into the active site of the *Hsa*NANA-9-PS model revealed several potential interactions between the enzyme and substrates. As mentioned in Section 6.2.1, the majority of the residues involved with substrate binding in *Nme*NANAS are conserved in the sequence of *Hsa*NANA-9-PS. Depicted in Figure 6.7B, the interactions of these conserved residues are shown to be maintained as well.

Interestingly, the *Hsa*NANA-9-PS docking experiment reveals several additional potential ligand interactions that were not identified through sequence comparison. Whilst the *Hsa*NANA-9-PS β2-α2 loop is rich in arginine and lysine (Figure 6.5), the only residue from the loop that is identified to hydrogen bond to the phosphate group of ManNAc-6-P is Thr79. It is possible, however, that while the identified arginine and lysine residues do not directly interact with the phosphate of ManNAc-6-P, they provide a positively charged environment that aids in substrate recognition. The docking results also predict the involvement of Lys290 in phosphate recognition. Lys290 is the only amino acid revealed by docking which is contributed by the opposite monomer and directly interacts with substrate.

It is notable that the *N*-acetyl group of ManNAc-6-P does not appear to interact with any amino acid side-chains. The interaction between the *N*-acetyl group of ManNAc and *Nme*Arg314 in *Nme*NANAS was shown to be essential to catalysis, and positions ManNAc for reaction with PEP (Chapter 2). It is possible that this interaction is not necessary in *Hsa*NANA-9-PS or that a similar mechanism remains as yet unidentified. Furthermore, the lack of a congruent residue to *Nme*Arg314 in *Hsa*NANA-9-PS is particularly interesting given that *Hsa*NANA-9-PS is able to catalyse the formation of 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid 9-phosphate (KDN-9-P) from PEP and mannose 6-phosphate (Man-6-P) (*182*). The ability of *Hsa*NANA-9-PS to catalyse KDN-9-P formation further indicates that the steering mechanism of *Nme*Arg314 is not required in *Hsa*NANA-9-PS. Conversely, *Nme*NANAS is unable to catalyse a reaction between Mannose and PEP, further implicating the necessity of the interaction between the N-acetyl group of ManNAc and *Nme*Arg314 in *Nme*NANAS.

6.3 Cloning, expression and purification

6.3.1 Cloning, expression and purification of *Hsa*NANA-9-PS

The *Hsa*NANA-9-PS construct was codon optimised for expression in *E. coli*, synthesised and cloned into pDONR221 by GeneArt®. Sub-cloning into pDEST17, expression and purification optimisation of *Hsa*NANA-9-PS was carried out by Thomas Cotton as outlined in Section 8.10.2

6.3.2 Cloning of *Nme*/*Hsa* chimeras

Two chimeric constructs were generated from the genes encoding *Nme*NANAS and *Hsa*NANA-9-PS using a multi-step PCR protocol that is outlined in Figure 6.8. The sequence encoding the linker and AFPL domain of *Nme*NANAS was stitched to the sequence encoding the *Hsa*NANA-9-PS catalytic domain to generate the *Hsa*/*Nme* chimera. The linker and AFPL domain of *Hsa*NANA-9-PS was stitched to the sequence encoding the *Nme*NANAS catalytic domain to generate the *Nme*/*Hsa* chimera. The template used for amplification of the *Nme*NANAS sequence was the pDEST17-*Nme*NANAS construct outlined in Chapter 2.

The position chosen for interchanging the domains was residue *Nme*Gly272, which is conserved in *Hsa*NANA-9PS as residue *Hsa*G274. This position is located prior to the linker sequence as indicated in Figure 6.2. The template used for amplification of the *Hsa*NANA-9-PS sequence was a pDEST17-*Hsa*NANA-9PS construct, sub-cloned from a sequence that was synthesised and codon optimised for expression in an *E. coli* system by GeneArt®. The methods used for PCR are outlined in Section 8.3.2, and primer sequences are displayed in Table 8.9.

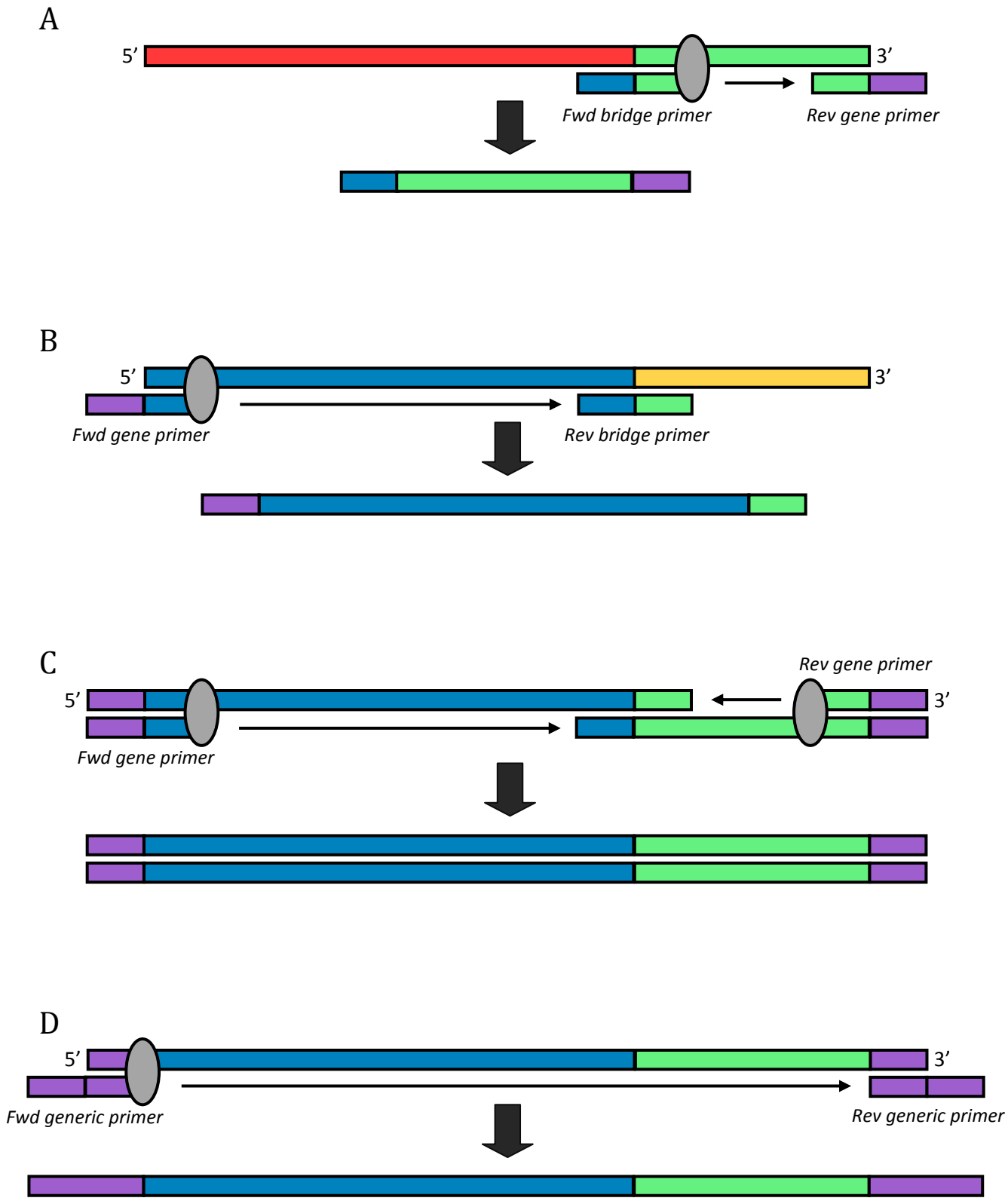


Figure 6.8 Assembly of NANA/NANA-9-PS chimera constructs. (A) Round 1: Amplification of AFPL domain from enzyme 1. (B) Round 2: Amplification of catalytic domain from enzyme 2. (C) Round 3: Fusion of amplification products from rounds 1 & 2. (D) Round 4: Extension of round 3 product using generic nested primers.

The first two steps of chimera generation involved the amplification of the catalytic domain and linker/AFPL domain from each donor sequence. Shown in Figure 6.8A, the linker/AFPL domain was amplified from the first enzyme's sequence using a gene specific reverse primer and forward bridging primer. The bridging primer comprised of sequence specific to the 5' terminus of the first enzyme's linker/AFPL domain and sequence specific to the 3' terminus of the second enzyme's catalytic domain. The resulting product was a linear construct of the first enzyme's linker/AFPL domain sequence with a 5' extension corresponding to sequence from the 3' terminus of the second enzyme's catalytic domain.

The amplification of the second enzyme's catalytic domain sequence is outlined in Figure 6.8B. The catalytic domain was amplified from the second enzyme's sequence using a gene specific forward primer and reverse bridging primer. The bridging primer consisted of sequence specific to the 5' terminus of the first enzyme's linker/AFPL domain and sequence specific to the 3' terminus of the second enzyme's catalytic domain. The resulting product was a linear construct of the second enzyme's catalytic domain sequence with a 3' extension corresponding to sequence from the 5' terminus of the second enzyme's linker/AFPL domain.

The third round of amplification involved stitching the linker/AFPL domain of the first enzyme to the catalytic domain of the second enzyme using overlap extension PCR (Figure 6.8C). The amplified products from rounds 1 and 2 were used as templates. Primers used for round 3 were the gene specific reverse primer from round 1 and the gene specific forward primer from round 2. The aforementioned gene specific nested primers were designed incorporating generic extensions for further rounds of nested PCR (Table 8.9). The product from the third round was the complete linear construct of the chimera with 5' and 3' extensions for nested PCR.

Following amplification of the linear chimera construct in round 3, the purified product was extended using generic gateway primers which incorporated sequence encoding an N-terminal TEV protease cleavage site (Figure 6.8D). Successful amplification at each stage was determined by agarose gel electrophoresis. Cloning and sub-cloning of the chimera constructs was achieved through the use of the gateway system as outlined in Section 8.3. The linear product from round 4 was cloned into the donor vector pDONR-221. Ligated product was subsequently transformed into *E. coli* One Shot® TOP10 cells (Invitrogen) as outlined in Section 8.3.10. Plasmids from transformants were purified and each of the chimera genes were successfully sequence verified in the donor vector. The chimera genes were sub-cloned into the destination vector pDEST-17, which encodes an N-terminal His-tag.

6.3.3 Expression and purification

The pDEST-17-*Nme*/*Hsa* chimera and pDEST-17-*Hsa*/*Nme* chimera were both transformed into three different expression cell lines to optimise the yield of soluble protein. The expression vectors were transformed into *E. coli* BL21 (DE3) Star cells, *E. coli* Rosetta™ 2 (DE3) cells and *E. coli* Bl21 (DE3) pBB540/pBB542 (Chaperone 3) cells. In order to compare induced against non-induced samples, two cultures of each cell line were grown overnight in LB, with one induced by IPTG and the other not. The results from the expression tests for the *Nme*/*Hsa* chimera and *Hsa*/*Nme* chimera were analysed by SDS-PAGE. The best yield of soluble protein for the *Nme*/*Hsa* chimera was attained using Chaperone 3 cells. Unfortunately, while the *Hsa*/*Nme* chimera did over-express with IPTG induction, there was no soluble protein observed in any of the expression systems used. In an effort to solubilise the *Hsa*/*Nme* chimera, the construct was sub-cloned into pDEST-15 which bears a GST solubility fusion tag and the expression test was repeated. Unfortunately, no soluble protein was afforded for the *Hsa*/*Nme* chimera.

The *Nme*/*Hsa* chimera was expressed and purified using a similar method to that used for *Nme*NANAS wild-type as outlined in Section 8.4. Samples of purified enzyme were analysed by SDS PAGE (Figure 6.9).

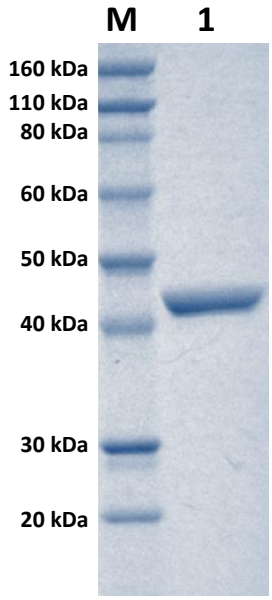


Figure 6.9 SDS-PAGE gel of purified *Nme*/*Hsa* chimera. Lane M contains marker with molecular weights defined. Lane 1 contains purified *Nme*/*Hsa* chimera.

6.4 Kinetic analysis of *Nme*/*Hsa* chimera

6.4.1 Attempted kinetic characterisation of *Nme*/*Hsa* chimeras

To determine the kinetic parameters of the *Nme*/*Hsa* chimera, the consumption of PEP was monitored using the assay described in Section 8.5.1. *Nme*/*Hsa* chimera was assayed with PEP and ManNAc/ManNAc-6-P at several different pH levels and temperatures as well as in the presence of a variety of metal ions. No kinetic activity could be observed for the *Nme*/*Hsa* chimera with either ManNAc of ManNAc-6-P even at high protein concentrations of approximately 1 mg/mL. Lack of activity equates to a k_{cat} less than 0.004 s^{-1}.

6.5 Binding studies of *Nme*/*Hsa* chimera

6.5.1 Isothermal titration calorimetry

Binding constants for Mn^{2+} and PEP were determined for the *Nme*/*Hsa* chimera using the methods described for *Nme*NANAS wild-type (Section 8.5.4). ITC results attained for the *Nme*/*Hsa* chimera are compared to previously determined dissociation constants (K_d) for *Nme*NANAS wild-type in Table 6.1 (*114*). *Nme*NANAS wild-type was used for comparison as the residues for Mn^{2+} and PEP binding are contributed solely by the catalytic domain and the *Nme*/*Hsa* chimera contains the *Nme*NANAS catalytic barrel. The raw experimental binding curve and one-site model fitted ITC data for Mn^{2+} and PEP with the *Nme*/*Hsa* chimera are displayed in Figure 6.10.

	***Nme*NANAS (*114*)**	***Nme*/*Hsa* Chimera**
K_d Mn^{2+} (μM)	10.6 ± 0.5	11.7 ± 0.4
K_d PEP (μM)	219 ± 11	374 ± 28

Table 6.1 ITC derived binding constants for wild-type *Nme*NANAS and the *Nme*/*Hsa* chimera.

To determine the K_d value for metal with the *Nme*/*Hsa* chimera, Mn^{2+} was titrated into metal-free enzyme. The Mn^{2+} affinity of the *Nme*/*Hsa* chimera was determined to be very similar to that measured for wild-type *Nme*NANAS. As depicted in Figure 6.10A, the binding of Mn^{2+} is an exothermic event in the *Nme*/*Hsa* chimera and is thus an enthalpically favourable process.

Shown in Table 6.1, the PEP affinity for the *Nme*/*Hsa* chimera is slightly lower than that observed for *Nme*NANAS wild-type. As shown in Figure 6.10B, the binding of PEP is an endothermic event in the *Nme*/*Hsa* chimera. This correlates with previous observations for PEP binding with wild-type *Nme*NANAS which was determined to be an entropically favourable event. The results from ITC indicate that the *Nme*/*Hsa* chimera behaves in a similar manner to that observed for *Nme*NANAS wild-type with respect to PEP and Mn^{2+} binding even though the chimera is not active (Section 6.4). These results indicate that the Mn^{2+} and PEP binding sites are still intact in the *Nme*/*Hsa* chimera.

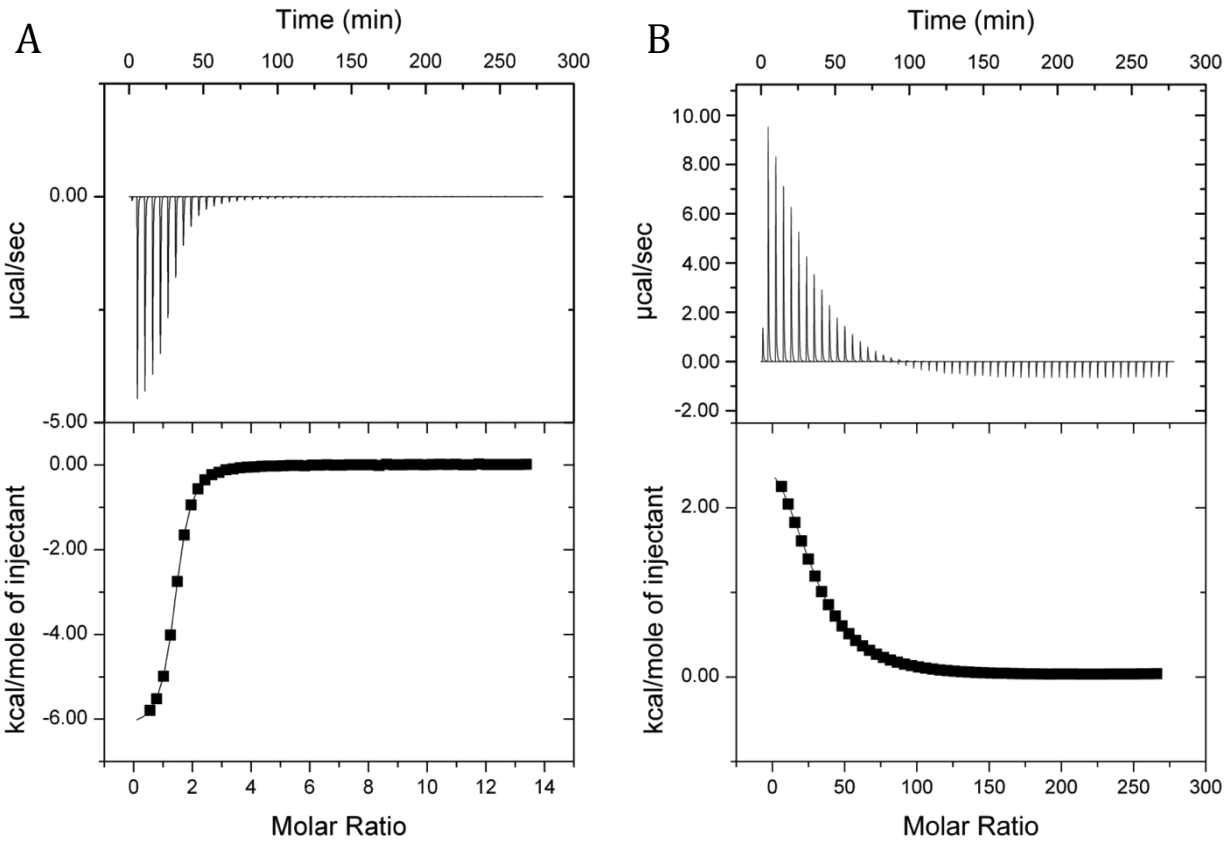


Figure 6.10 Enthalpy changes measured by ITC. (A) Mn^{2+} (5 mM) titrated into metal-free *Nme*/*Hsa* chimera (65 µM). (B) PEP (50 mM) titrated into *Nme*/*Hsa* chimera (40 µM) in the presence of Mn^{2+} (1 mM).

6.5.2 Thermal stability

The effects of Mn^{2+}, PEP and ManNAc/rManNAc/ManNAc-6-P on the thermal stability of the *Nme*/*Hsa* chimera were assessed by DSF. The protocol used for DSF is outlined in Section 8.5.3. Protein used for DSF was treated with 10 mM EDTA prior to SEC to ensure removal of trace metals. The melting points determined for each sample were plotted graphically and compared to data obtained for *Nme*NANAS and *Hsa*NANA-9-PS wild-type enzymes (Figure 6.11 and Table 6.2).

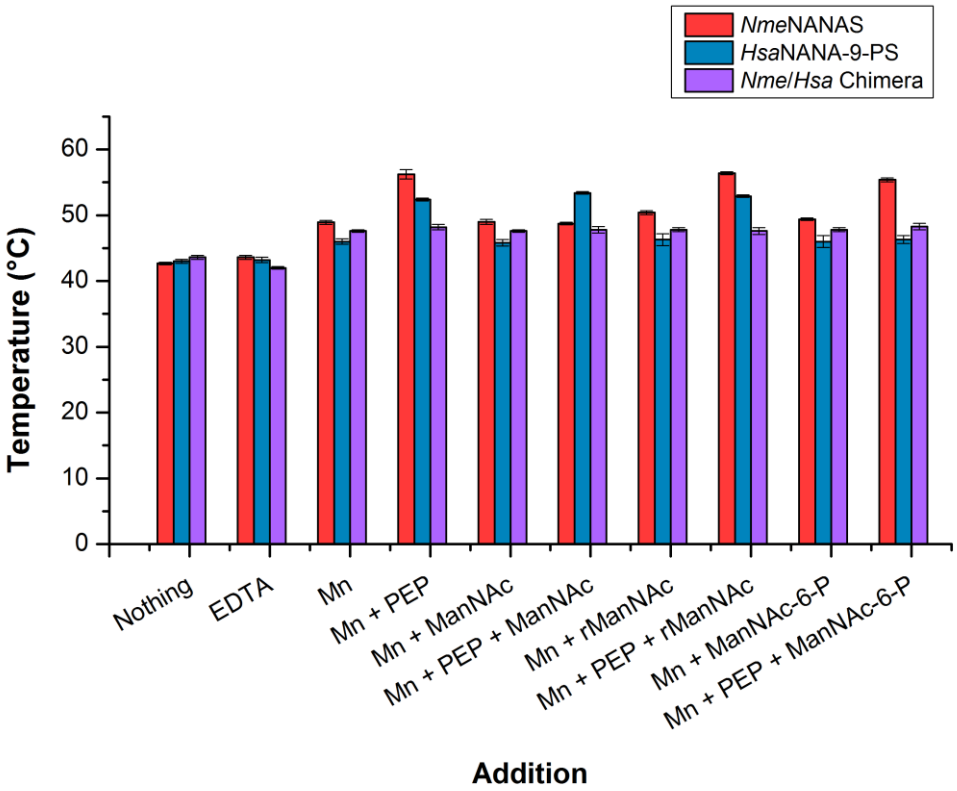


Figure 6.11 Bar graph of thermal melting points for *Nme*NANAS wild-type (red), *Hsa*NANA-9-PS wild-type (blue) and *Nme*/*Hsa* chimera (purple) in the presence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc/ManNAc-6-P at 10 mM. ManNAc/rManNAc/ManNAc-6-P was used at 1 mM when in combination with PEP.

Addition	*Nme*NANAS (°C)	*Hsa*NANAS (°C)	*Nme*/*Hsa* Chimera (°C)
Nothing	42.7 ± 0.2	43.0 ± 0.3	43.6 ± 0.3
EDTA	43.6 ± 0.3	43.2 ± 0.4	42.0 ± 0.2
Mn^{2+}	48.9 ± 0.3	46.0 ± 0.4	47.6 ± 0.2
Mn^{2+} + PEP	56.2 ± 0.7	52.4 ± 0.2	48.2 ± 0.4
Mn^{2+} + ManNAc	49.0 ± 0.4	45.8 ± 0.5	47.6 ± 0.2
Mn^{2+} + PEP + ManNAc	48.7 ± 0.2	53.4 ± 0.2	47.8 ± 0.5
Mn^{2+} + rManNAc	50.4 ± 0.3	46.3 ± 0.9	47.8 ± 0.3
Mn^{2+} + PEP + rManNAc	56.4 ± 0.2	52.9 ± 0.2	47.6 ± 0.5

Table 6.2 Thermal melting points for *Nme*NANAS wild-type, *Hsa*NANA-9-PS wild-type and *Nme*/*Hsa* chimera in the presence of ligand as determined by DSF. EDTA, Mn^{2+} and PEP were each used at a concentration of 1 mM and ManNAc/rManNAc/ManNAc-6-P at 10 mM. ManNAc/rManNAc/ManNAc-6-P was used at 1 mM when in combination with PEP.

Apoenzyme *Nme*/*Hsa* chimera was found to melt at 43.6 °C (Figure 6.11 and Table 6.2). The melting point of the *apo*-*Nme*/*Hsa* chimera is very similar to that for both *Nme*NANAS and *Hsa*NANA-9-PS wild-type enzymes which melt at 42.7 ± 0.2°C and 43.0 ± 0.3°C respectively. The addition of 1 mM EDTA to *apo*- *Nme*/*Hsa* chimera did not influence the thermal stability of the enzyme which likely indicates the successful removal of trace metal ions during purification. The addition of 1 mM Mn^{2+} increased the melting point of *Nme*/*Hsa* chimera, by approximately 4 °C. Shown in Figure 6.11 and Table 6.2, Mn^{2+} stabilises both *Nme*NANAS and *Hsa*NANA-9-PS wild-type enzymes also.

As discussed in Chapter 3, metal bound *Nme*NANAS is significantly stabilised upon the addition of PEP. Shown in Figure 6.11 and Table 6.2, *Hsa*NANA-9-PS is similarly stabilised by PEP. Interestingly, the *Nme*/*Hsa* chimera is not thermally stabilised in the presence of PEP. This is surprising given that the *Nme*/*Hsa* chimera is shown to bind both metal and PEP in a similar fashion to *Nme*NANAS wild-type as discussed in Section 6.5.1. Furthermore, the *Nme*/*Hsa* chimera is not thermally stabilised by any of the combinations of PEP with ManNAc, rManNAc or ManNAc-6-P.

6.6 Structural analysis of *Nme*/*Hsa* chimera

6.6.1 Circular dichroism

CD was used to assess and compare the secondary structure of *Nme*/*Hsa* chimera with *Nme*NANAS and *Hsa*NANA-9-PS wild-types (Figure 6.12). The K2D3 server (*142*) was used to estimate the secondary structure composition from CD spectra. Shown in Table 6.3, estimated secondary structure is compared to values derived from the crystal structure of *Nme*NANAS (PDB code: 1XUZ) and *Hsa*NANA-9-PS model.

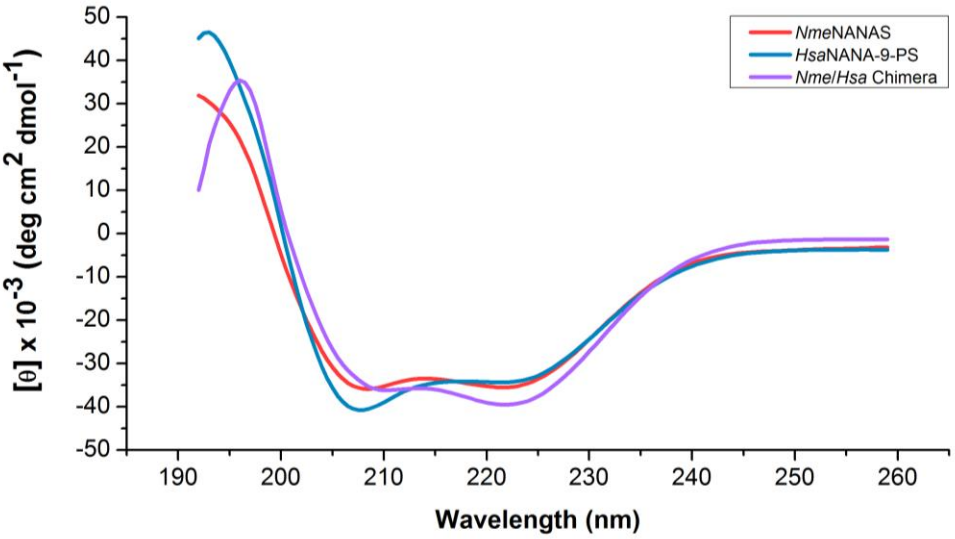


Figure 6.12 CD spectra for *Nme*NANAS (red), *Hsa*NANA-9-PS (blue) and *Nme*/*Hsa* chimera (purple).

	***Nme*NANAS**	***Nme*NANAS (structure)**	***Hsa*NANA-9-PS**	***Hsa*NANA-9-PS (model)**	***Nme*/*Hsa* Chimera**
α-helix	59.9 %	50.3 %	54.3 %	50.1 %	54.9 %
β-sheet	12.3 %	15.5 %	15.2 %	15.8 %	13.8 %
other	27.8 %	34.2 %	30.5 %	34.1 %	31.3 %

Table 6.3 Predictions of secondary structure composition for *Nme*NANAS, *Hsa*NANA-9-PS and *Nme*/*Hsa* chimera calculated from CD data using K2D3 server (*142*). Theoretical secondary structure composition of *Nme*NANAS and *Hsa*NANA-9-PS calculated from crystallographic coordinates 1XUZ and *Hsa*NANA-9-PS model respectively.

Since kinetic activity of the *Nme*/*Hsa* chimera could not be detected (Section 6.4), CD was used to ascertain whether the fusion protein was folded in solution. As depicted in Figure 6.12 and Table 6.3, the *Nme*/*Hsa* chimera is folded in solution and has comparable levels of α-helix and β-sheet to both *Nme*NANAS and *Hsa*NANA-9-PS. Depicted in Figure 6.12, the CD profile of the *Nme*/*Hsa* chimera is slightly different to that of *Nme*NANAS and *Hsa*NANA-9-PS.

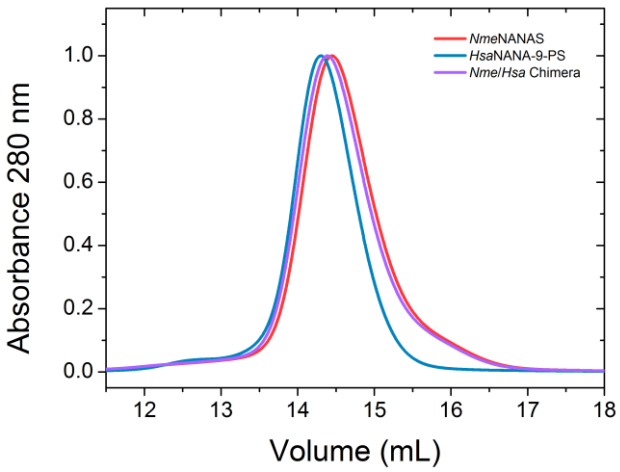
6.6.2 Mass spectrometry

The molecular mass of His_6-tagged *Nme*/*Hsa* chimera was analysed using MS. A theoretical mass of 42,930 Da was calculated for the *Nme*/*Hsa* chimera from its peptide sequence using ProtParam (*141*). The experimental mass of the *Nme*/*Hsa* chimera as determined by MS was 42,926 Da which differs by only 0.01 % from the aforementioned calculated mass.

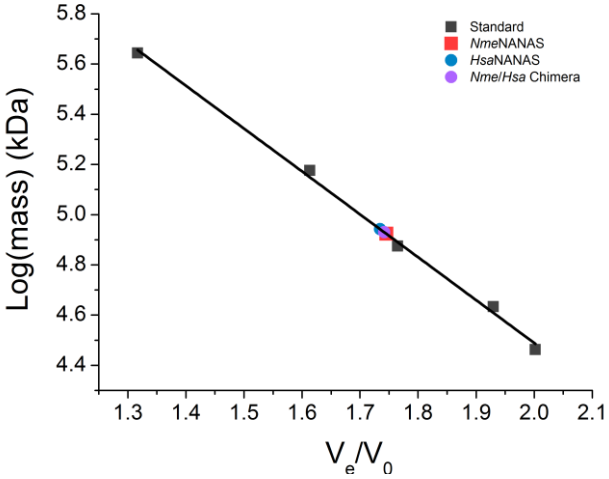
6.6.3 Analytical size exclusion chromatography

The quaternary state of the *Nme*/*Hsa* chimera was investigated via analytical SEC using the method outlined in Section 8.5.6. A calibration curve was generated using protein standards of known molecular weight. The elution volume (V_e) of protein standards relative to the column void volume (V_0) was plotted against the log of the protein standards molecular mass. A linear fit of the analytical SEC calibration data was determined to have an R^2 value of 0.987. The calibration curve was subsequently used to extrapolate the molecular masses of each of the *Nme*/*Hsa* chimera as shown in Figure 6.13.

A



B



C

	***Nme*NANAS**	***Hsa*NANA-9-PS**	***Nme*/*Hsa* Chimera**
V_e/V_0	1.74	1.73	1.74
Estimated mass (kDa)	81	85	83
Theoretical mass of dimer (kDa)	77	81	79

Figure 6.13 (A) Normalised analytical SEC chromatograms for *Nme*NANAS (red), *Hsa*NANA-9-PS (blue) and *Nme*/*Hsa* chimera (purple). (B) Analytical SEC calibration curve. Protein standards are marked in grey, *Nme*NANAS in red, *Hsa*NANA-9-PS in blue and *Nme*/*Hsa* chimera in purple. A line of best fit is plotted in black. (C) Table of estimated and theoretical masses for *Nme*NANAS and *Hsa*NANA-9-PS and the *Nme*/*Hsa* chimera.

As shown in Figure 6.13, the *Nme*/*Hsa* chimera elutes as a single species at a molecular weight corresponding to 83 kDa. The molecular weight of *Nme*/*Hsa* chimera determined from analytical SEC is thus similar to the theoretical mass for dimeric *Nme*/*Hsa* chimera. Displayed in Figure 6.13C, both *Nme*NANAS and *Hsa*NANA-9-PS elute as dimeric enzymes also. The results from analytical SEC indicate that the *Nme*/*Hsa* chimera is a dimer in solution.

6.6.4 Attempted crystallisation

Several attempts at crystallising the *Nme*/*Hsa* chimera were made. The *Nme*/*Hsa* chimera was screened using conditions modified from those previously described for crystallising wild-type *Nme*NANAS (*112*). Protein crystallisation conditions were also screened using commercially available Morpheus, PACT and JCSG screens. Protein was concentrated to 6 mg/mL and 15 mg/mL and screened in the absence and presence of ligands. Screens were incubated at 5 °C and 20 °C for several months, but none of the conditions that were screened yielded crystals of diffraction quality.

6.6.5 Small-angle X-ray scattering

Given the lack of success in crystallising the *Nme*/*Hsa* chimera, SAXS was used to assess the chimera's solution structure. The *Nme*/*Hsa* chimera was analysed in the presence of 1 mM Mn^{2+} and 1 mM PEP. The SAXS profile of the *Nme*/*Hsa* chimera was determined using the methods outlined in Sections 8.5.8 and 8.5.9. Displayed in Figure 6.14A, the scattering profile of the *Nme*/*Hsa* chimera is compared to the profiles of *Nme*NANAS and *Hsa*NANA-9-PS in the presence of ligand. Since a crystal structure of the *Nme*/*Hsa* chimera could not be obtained, the experimental scattering of the *Nme*/*Hsa* chimera was fitted to CRYSOL (*153*) plots generated from the *Nme*NANAS crystal structure 1XUZ (*112*) and the *Hsa*NANA-9-PS homology model (Figure 6.14B).

	***Nme*NANAS**	***Hsa*NANA-9-PS**	***Nme*/*Hsa* Chimera**
I_0 (cm^{-1})	0.082 ± 0.001	0.079 ± 0.001	0.076 ± 0.001
R_g (Å)	33.0 ± 0.5	35.1 ± 0.8	33.6 ± 0.1
D_{max} (Å)	116	123	118
Porod volume ($Å^3$)	121,724	125,364	123,358
Estimated mass (kDa)	85	88	86
Theoretical mass of dimer (kDa)	77	81	79

Table 6.4 SAXS structural parameters determined for *Nme*/*Hsa* chimera compared with parameters for *Nme*NANAS and *Hsa*NANA-9-PS. SAXS data obtained at a wavelength of 1.0332 Å.

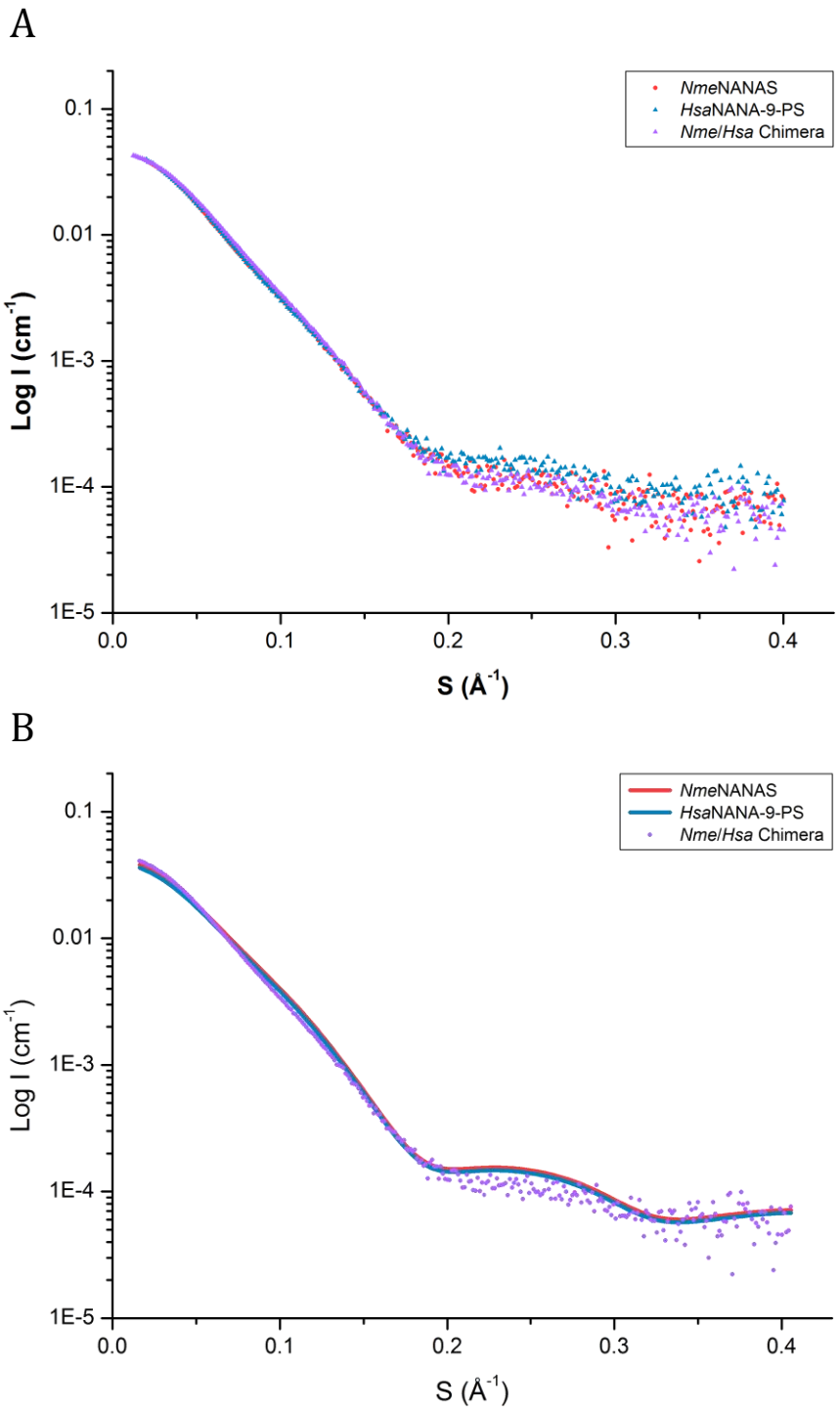


Figure 6.14 SAXS profile of *Nme*/*Hsa* chimera in the presence of 1 mM Mn^{2+} and 1 mM PEP. (A) Comparison of SAXS data for *Nme*/*Hsa* chimera (purple) with *Nme*NANAS wild-type (red) and *Hsa*NANA-9-PS wild-type (blue). (B) CRYSOL fits of *Nme*/*Hsa* chimera SAXS data with theoretical scattering of *Nme*NANAS wild-type generated from crystallographic coordinates 1XUZ (red line) and theoretical scattering generated from the *Hsa*NANA-9-PS wild-type homology model (blue line). SAXS data was collected using approximately 10 mg mL^{-1} protein in 30 mM triethanolamine-HCl buffer at pH 7.5. Samples were eluted by SEC (Superdex 200 5/150).

The *Nme*/*Hsa* chimera structure parameters obtained from SAXS are comparable to those obtained for both *Nme*NANAS and *Hsa*NANA-9-PS wild-types (Table 6.4). As depicted in Figure 6.14A, the scattering profile of the *Nme*/*Hsa* chimera is also similar to those attained for *Nme*NANAS and *Hsa*NANA-9-PS. The molecular mass of liganded *Nme*/*Hsa* chimera was calculated from SAXS data using SAXS MoW (*154*), and was estimated to be 86 kDa, which is similar to that expected for dimeric protein (Table 6.4). The results from SAXS comply with data obtained from analytical SEC which also indicate that the *Nme*/*Hsa* chimera is dimeric (Section 6.6.3).

Shown in Figure 6.14B, the SAXS data obtained for the *Nme*/*Hsa* chimera fits well to the theoretical scattering profile generated from the liganded *Nme*NANAS crystal structure (Chi value of 1.56). The fit to the theoretical scattering profile of the *Hsa*NANA-9-PS model is also good with a Chi value of 1.58. The similar theoretical scattering profiles generated for *Nme*NANAS and *Hsa*NANA-9-PS are likely caused by their high structural similarity. This similarity is forced by the method of model construction where the *Nme*NANAS structure was used as a template for the *Hsa*NANA-9-PS model (Section 6.2.2). The overall structure of the *Hsa*NANA-9-PS model is thus very comparable to that of the *Nme*NANAS crystal structure and the resulting theoretical scattering profiles are almost identical.

Described in Chapter 3, wild-type *Nme*NANAS undergoes a conformational change in the presence of Mn^{2+} and PEP that is detectable by SAXS. The conformational change in *Nme*NANAS is linked to a hydrogen bonding network that involves the rearrangement of the ManNAc binding site upon liganding of PEP. Shown in Figure 6.15, the SAXS profile of the *Nme*/*Hsa* chimera resembles the profile of liganded and not ligand-free *Nme*NANAS. This may indicate that the conformational change induced by PEP binding is still possible in the *Nme*/*Hsa* chimera.

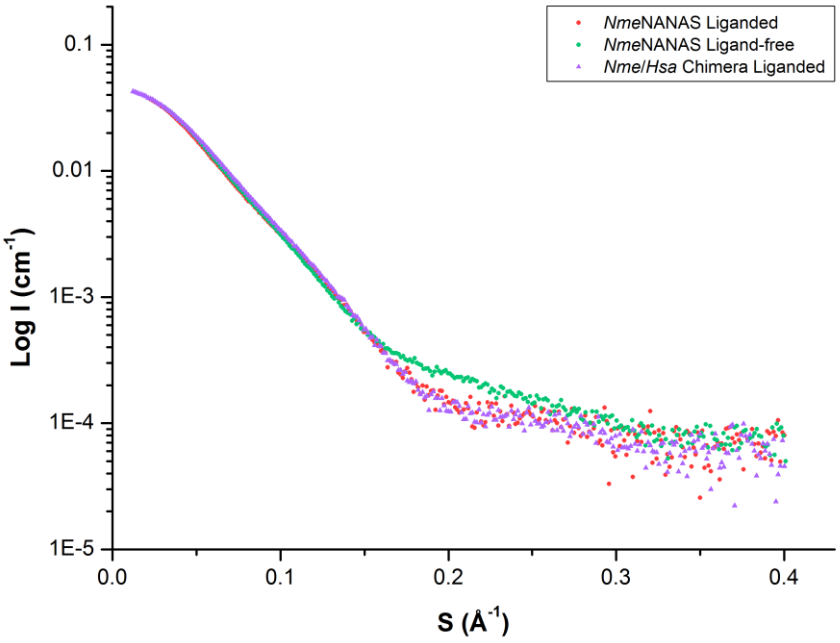


Figure 6.15 SAXS profile of *Nme*/*Hsa* chimera in the presence of 1 mM Mn^{2+} and 1 mM PEP compared with liganded and ligand-free scattering for *Nme*NANAS wild-type. SAXS data for liganded *Nme*/*Hsa* chimera is in purple, liganded *Nme*NANAS wild-type in red and ligand-free *Nme*NANAS wild-type in green.

Residue *Nme*Glu282 was identified to play an important role in the hydrogen bonding network of *Nme*NANAS (Chapter 3). Shown in Figure 6.2, *Nme*Glu282 is conserved as *Hsa*Glu284. MD simulations outlined in Chapter 3 revealed that *Nme*Glu282 forms a hydrogen bond with *Nme*Arg194 of the opposing monomer. In the *Nme*/*Hsa* chimera, residues Gly272 onwards are contributed by the *Hsa*NANA-9-PS sequence. An alignment of the *Hsa*NANA-9-PS model with *Nme*NANAS crystal structure 1XUZ (*112*) is depicted in Figure 6.16, where it can be observed that residue *Hsa*Glu284 would still be in a suitable position to interact with *Nme*Arg194 in the *Nme*/*Hsa* chimera.

The *Nme*Glu282-*Nme*Arg194 is the only interaction from the hydrogen bonding network identified in Chapter 3 that is retained in the *Nme*/*Hsa* chimera. Interestingly, the *Nme*Glu282-*Nme*Arg194 interaction is not retained in the *Hsa*/*Nme* chimera since *Nme*Arg194 is not conserved in *Hsa*NANA-9-PS. *Nme*Arg194 is instead replaced by *Hsa*Asn196 (Figure 6.2) which may not have a long enough side-chain to hydrogen bond with *Nme*Glu282. The abolition of the entire hydrogen bonding network in the *Hsa*/*Nme* chimera may be the cause of its insolubility. It is also possible that insolubility of the *Hsa*/*Nme* chimera is caused by repulsive electrostatic interactions not present in the *Nme*/*Hsa* chimera, or an inability to bury hydrophobic surfaces resulting in dimer instability.

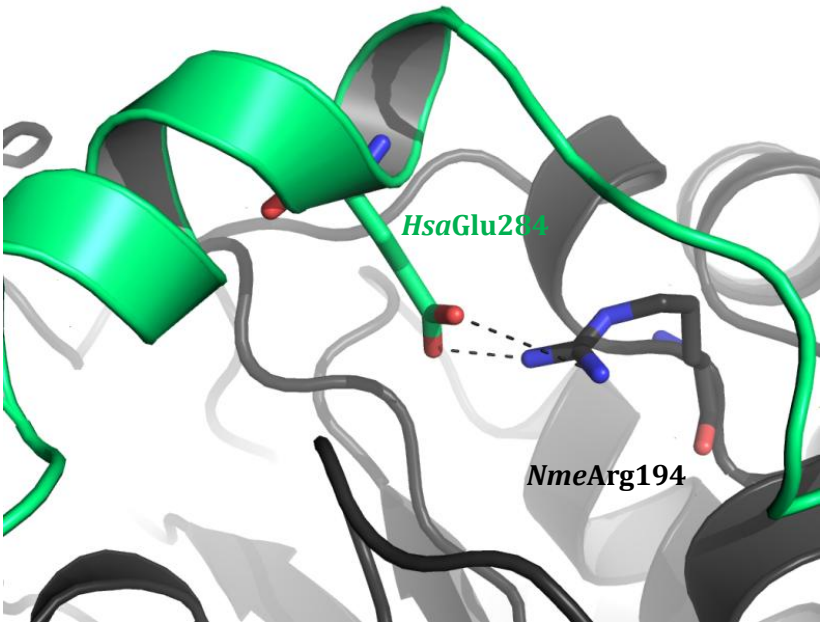


Figure 6.16 Alignment of *Nme*NANAS crystal structure (catalytic domain in grey) with *Hsa*NANA-9-PS model (linker and AFPL domain in green) showing the possible H-bond interaction (black dashes) between *Nme*Arg194 and *Hsa*Glu284 in the *Nme*/*Hsa* chimera.

6.6.6 Analytical Ultracentrifugation

To determine the stability of the *Nme*/*Hsa* chimera quaternary state, sedimentation velocity experiments were caried out at varied protein concentrations. To determine if the quaternary state of the *Nme*/*Hsa* chimera was concentration dependent, samples were run at 0.1 mg/mL, 0.5 mg/mL and 1.0 mg/mL. Displayed in Figure 6.17, data was fitted to a continuous size-distribution (c_s) model using the program SEDFIT (*150, 151*). Weight average molecular masses were calculated using weight average sedimentation coefficients and f/f_0 values from c_s fits.

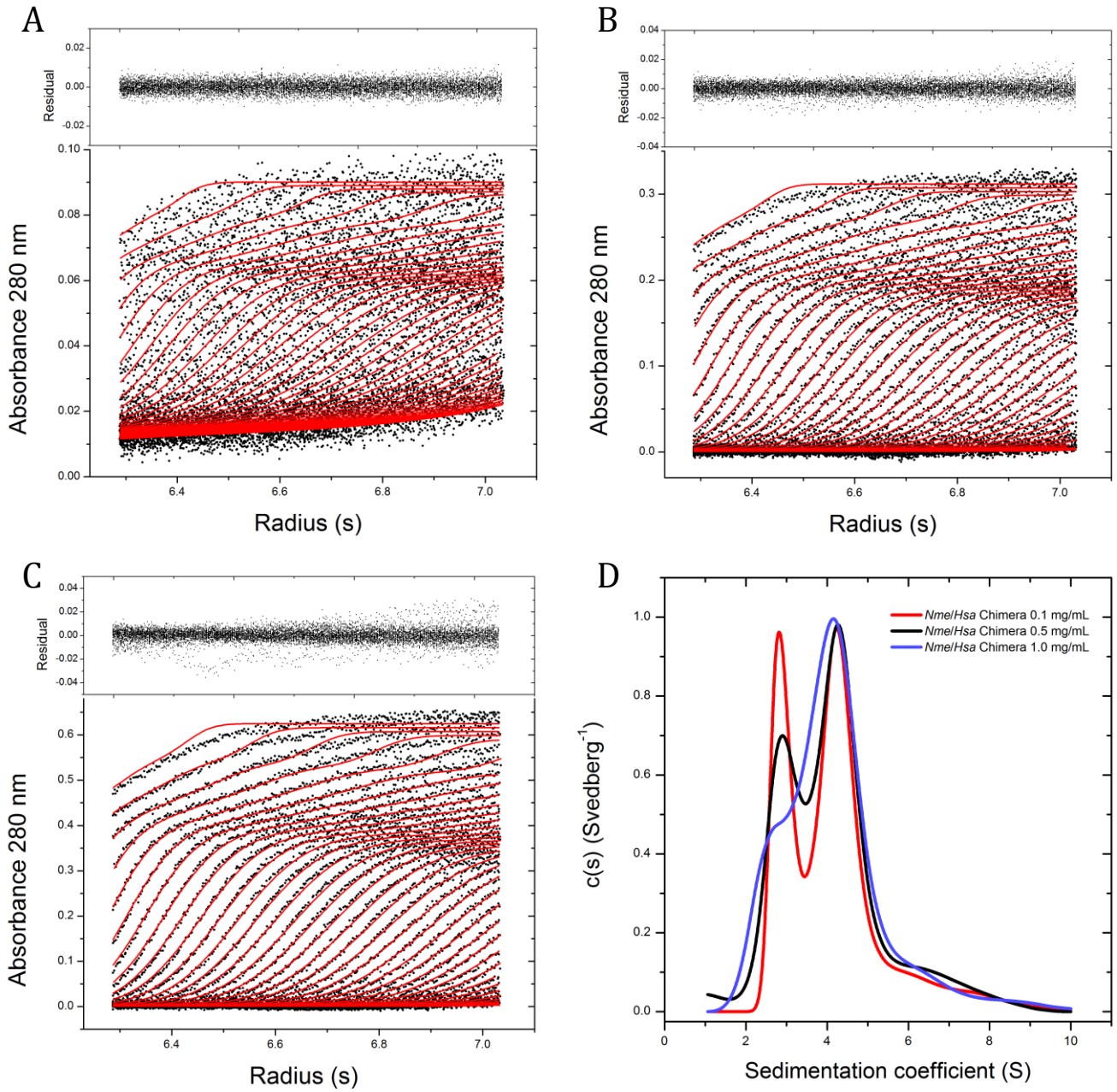


Figure 6.17 Analytical ultra centrifuge fits measured at 280 nm absorbance. (A) *Nme*/*Hsa* chimera at 0.1 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (B) *Nme*/*Hsa* chimera at 0.5 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (C) *Nme*/*Hsa* chimera at 1.0 mg/mL, residual data (top), and fit of model to data from SEDFIT (bottom). (D) Normalised c(s) profiles for *Nme*/*Hsa* chimera at 0.1 mg/mL (red), 0.5 mg/mL (black) and 1.0 mg/mL (purple).

Displayed in Figure 6.17, the *Nme*/*Hsa* chimera was revealed to exist as two species at the concentrations used for AUC. In Figure 6.17D, there is a peak corresponding to an estimated molecular weight of 38 kDa (close to that calculated for monomer *Nme*/*Hsa* chimera) and a second peak to 81 kDa (close to that calculated for dimer *Nme*/*Hsa* chimera). A concentration dependent equilibrium is observed between these two species where the first peak is increasingly prominent at lower protein concentrations and the second peak is increasingly prominent at higher protein concentrations. The results from AUC suggest that at lower concentrations the *Nme*/*Hsa* chimera favours a monomeric quaternary state whereas at higher concentrations a dimeric state is favoured. Both *Nme*NANAS and *Hsa*NANA-9-PS are confirmed to be stable dimers indicating that the interactions that stabilise their respective quaternary states have been disrupted in the *Nme*/*Hsa* chimera. Furthermore, the c_s profiles displayed in Figure 6.17D reveal that there is minor protein aggregation present at higher sedimentation coefficient values.

Described in Sections 6.5.3 and 6.5.5, the results from both analytical SEC and SAXS imply that the *Nme*/*Hsa* chimera exists as a dimer. However, the protein concentrations used for both analytical SEC and SAXS were much greater than those used for AUC. As revealed by AUC, the *Nme*/*Hsa* chimera is primarily dimeric at 1 mg/mL. Consequentially, experiments done above this concentration would suggest that the chimera is dimeric.

Similarly, the results from ITC were obtained at protein concentrations where the chimera is likely to have been dimeric whereas for DSF the chimera is likely to have been monomeric. Outlined in Chapter 3, the conformational change of *Nme*NANAS imposed by PEP binding relies on the hydrogen bonding network that links the AFPL domain to the catalytic domain. The hydrogen bonding network is dependent on a dimeric quaternary state as the key interactions are contributed between opposing chains. Without the domain crossed homodimeric structure, the hydrogen bonding network is defunct. The ability of the *Nme*/*Hsa* chimera to undergo conformational change in the presence of PEP, could thus be observed by SAXS, where the *Nme*/*Hsa* chimera is dimeric. Conversely, the thermal stabilisation of the protein which is associated with PEP binding, could not be observed by DSF, where the *Nme*/*Hsa* chimera is monomeric.

6.7 Summary of results

The *Hsa*/*Nme* and *Nme*/*Hsa* chimeras were each successfully cloned from the genes encoding *Nme*NANAS and *Hsa*NANA-9-PS. The *Nme*/*Hsa* chimera was solubly expressed however soluble expression of the *Hsa*/*Nme* chimera could not be coerced even with the fusion of an N-terminal GST-tag. The *Nme*/*Hsa* chimera was successfully purified by metal affinity and size exclusion chromatography. Unfortunately, no activity could be observed with either ManNAc or ManNAc-6-P even at high enzyme concentrations.

Although the *Nme*/*Hsa* chimera was inactive, the protein was still able to bind both Mn^{2+} and PEP with dissociation constants comparable to that of wild-type *Nme*NANAS. The *Nme*/*Hsa* chimera did not share a similar thermal melting profile with either wild-type *Nme*NANAS or *Hsa*NANA-9-PS. As determined by DSF, *apo*-*Nme*/*Hsa* chimera melted at a similar temperature to both *apo*-*Nme*NANAS and *apo*-*Hsa*NANA-9-PS and was similarly stabilised by Mn^{2+}. Unlike the parent enzymes however, the *Nme*/*Hsa* chimera was not significantly stabilised in the presence of both PEP and Mn^{2+}.

Results obtained from SAXS indicate that the *Nme*/*Hsa* chimera is capable of undergoing a similar conformational change to that observed for wild-type *Nme*NANAS and *Hsa*NANA-9-PS in the presence of PEP and Mn^{2+}. Outlined in Chapter 3, a hydrogen bonding network links the catalytic and AFPL domains which are reorganised by PEP liganding to bind ManNAc. Only the *Nme*Glu282-*Nme*Arg194 interaction from the hydrogen bonding network is retained in the *Nme*/*Hsa* chimera through the conservation of residue *Hsa*Glu284.

The molecular mass of the *Nme*/*Hsa* chimera was successfully ascertained using MS and the protein was determined to be folded in solution using CD. At high protein concentrations (above 1 mg/mL), the *Nme*/*Hsa* chimera was revealed to be a dimer by SAXS, Analytical SEC and AUC. AUC also uncovered a concentration dependent quaternary state of the *Nme*/*Hsa* chimera at protein concentrations below approximately 1 mg/mL.

The results from AUC imply that under the conditions used for SAXS and ITC, the *Nme*/*Hsa* chimera was dimeric and for DSF the *Nme*/*Hsa* chimera was monomeric. This correlates with observations that PEP did not stabilise the *Nme*/*Hsa* chimera in DSF but had similar ITC and SAXS profiles to wild-type *Nme*NANAS with PEP.

6.8 Discussion

The work described in Chapter 6 sought to interconvert the catalytic functions of a bacterial NANAS with a mammalian NANA-9-PS. This was to be achieved by interchanging the AFPL domains of *Nme*NANAS and *Hsa*NANA-9-PS to create two chimeric proteins; one with an *Nme*NANAS catalytic domain joined to a *Hsa*NANA-9-PS linker and AFPL domain (*Nme*/*Hsa* chimera), and one with a *Hsa*NANA-9-PS catalytic domain joined to an *Nme*NANAS linker and AFPL domain (*Hsa*/*Nme* chimera).

The *Hsa*/*Nme* and *Nme*/*Hsa* chimeras were both successfully cloned, however, only the *Nme*/*Hsa* chimera was solubly expressed. Unfortunately, neither NANAS nor NANA-9-PS activity was observed with the *Nme*/*Hsa* chimera. Though inactive, DSF and ITC indicated that the *Nme*/*Hsa* chimera was able to bind both Mn^{2+} and PEP. Unlike *Nme*NANAS and *Hsa*NANA-9-PS, the chimeric protein was not thermally stabilised in the presence of PEP. SAXS analysis indicates that the *Nme*/*Hsa* chimera retains the ManNAc binding-site pre-organisation mechanism of *Nme*NANAS*,* however, these results conflict with the lack of thermal stabilisation observed via DSF.

AUC revealed that at high protein concentrations (like those used for ITC and SAXS), the *Nme*/*Hsa* chimera was dimeric, whilst at low protein concentrations (like those used for DSF), the *Nme*/*Hsa* chimera was monomeric. The concentration dependent quaternary state of the *Nme*/*Hsa* chimera is the likely cause for the variable PEP binding capability observed, as the protein would need to be dimeric for the hydrogen bonding relay to function.

Chapter 7

Discussion

7.1 Chapter outline

This chapter summarises and interprets the key findings made in this thesis. The role of the AFPL domain is discussed in detail, as are the evolutionary relationships of the SAS family. Potential, future analyses which continue the research outlined in this thesis are also discussed. The work described in this thesis builds upon previous studies of the SASs (*5, 112, 113*); enzymes which are responsible for the biosynthesis of sialic acids. Sialic acids are involved in many important biological processes, such as cell signalling, cell motility and evasion of host immune systems. The investigation of the enzymes responsible for sialic acid synthesis provides valuable insight into their numerous roles.

7.2 Role of the AFPL domain

7.2.1 Contribution of a specific catalytic residue

$(\beta/\alpha)_8$ Barrel extensions serve many roles in nature (*183-186*). They have been observed to have functional roles such as aiding in catalysis or regulation (*183, 185*) and have also been shown to provide structural support (*184*). The AFPL domain of the SAS family is a unique $(\beta/\alpha)_8$ barrel decoration (*112*). Described in Chapters 4 and 5 of this thesis, sequence pertaining to the AFPL domain is observed in both mammalian and bacterial SASs, with the only exception being members of the bacterial Type IV SAS clade. The crystal structure of *Nme*NANAS revealed that the AFPL domain contributes a single arginine residue into the active site of the opposing monomer (Figure 7.1). Residue ^{Nme}Arg314 contributes hydrogen bonding interactions with the *N*-acetyl functionality of rManNAc and hence ManNAc. ^{Nme}Arg314 is completely conserved in bacterial SASs which contain an AFPL domain. The high conservation of this arginine residue implicates its importance in bacterial SASs.

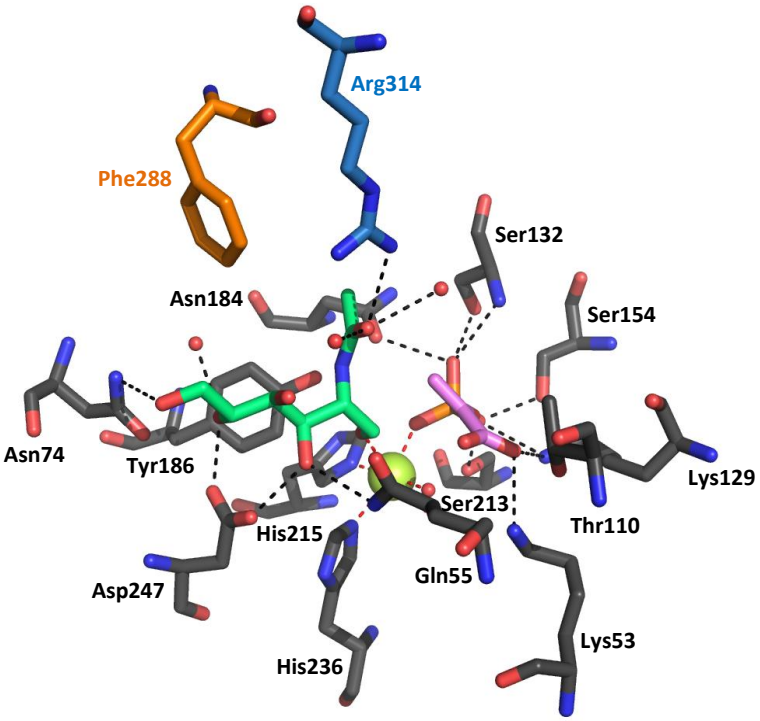


Figure 7.1 Close up of *Nme*NANAS active site. Residues contributed by the catalytic domain are in grey, linker in orange and AFPL domain in blue. ManNAc is coloured green, PEP pink, Mn^{2+} yellow and waters are shown in red. H-bonds are depicted with black dashes and dipolar bonds with red dashes. *Nme*NANAS PDB coordinates (1XUZ) (*112*).

From the crystal structure of *Nme*NANAS, the role of ^{Nme}Arg314 appears to be in the binding of ManNAc, however, the results obtained in Chapter 2 of this thesis indicate that the guanidinium functionality of ^{Nme}Arg314 is vital in steering ManNAc into a reactive position with PEP. Whilst removal of the functionality of ^{Nme}Arg314 abolishes activity, it does not remove the ability of the enzyme to bind ManNAc, implying that there may be other, more vital interactions involved in binding of the NANAS sugar substrate.

As previously mentioned, ^{Nme}Arg314 is fully conserved in bacterial SASs which bear AFPL domains. The structure of *Ssu*SAS, detailed in Chapter 5, similarly positions the congruent catalytic Arg residue (^{Ssu}Arg301), into the active site of the opposing monomer. The conservation of the ^{Nme}Arg314 residue in LegSs and PseSs is not surprising given that the sugar substrate of each clade contains an *N*-acetyl functional group at position C-2 (Figure 7.2). The C-2 *N*-acetyl group of the Pse sugar substrate, 2,4-diacetamido-2,4,6-trideoxyaltrose, has inverted stereochemistry when compared with the sugar substrates of NANAS and LegS. It is unknown if or how the difference in substrate stereochemistry would affect the role of the catalytic Arg in PseSs, however, the full conservation of ^{Nme}Arg314 in the PseS clade implies that the residue plays an important role. It is highly likely that the role of the conserved Arg residue in NANASs, LegSs and PseSs is similar to that observed for *Nme*NANAS, and that the charged guanidinium group steers each of these substrates into a reactive position.

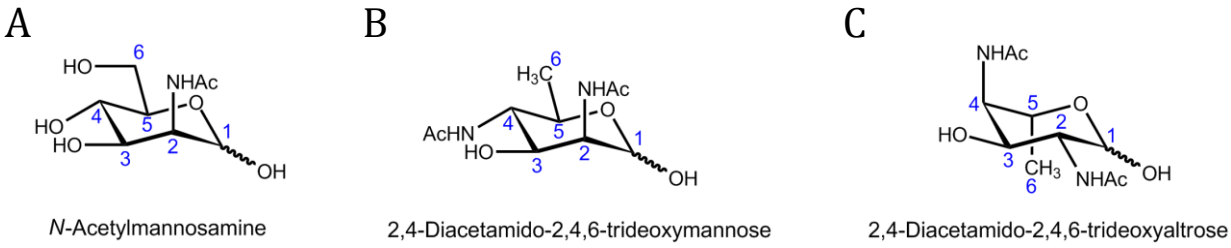


Figure 7.2 The sugar substrates of (A) NANAS, (B) LegS and (C) PseS.

The lack of an AFPL domain and thus a corresponding residue to ^{Nme}Arg314 in Type IV SASs, poses an interesting question as to how the enzymes of this clade position their aldehyde substrate for reaction. Given that the aldehyde substrate of *Cte*SAS could not be determined, it is not possible to make any precise inferences into this query. The absence of the AFPL domain and catalytic Arg in Type IV SASs may give some insight into the identity of the sugar substrate for this clade. It is possible that the catalytic Arg is not required in Type IV SASs because the aldehyde substrate of the Type IV clade does not have an *N*-acetyl functionality at position C-2. The simplest 6 carbon ManNAc derivative with this parameter – mannose, was assayed with *Cte*SAS with no observable activity. Further investigation is required to uncover the substrate and mechanism of this enzyme.

Although mammalian NANA-9-PSs also bear AFPL domains (*115*), ^{Nme}Arg314 is absent from the somewhat limited, available sequences of mammalian SASs. Instead, a fully conserved Val residue is present in each of the known mammalian SAS sequences. The absence of a congruent residue to ^{Nme}Arg314 in mammalian SASs is particularly interesting given that the substrate of NANA-9-PS is ManNAc-6-P - which, like the substrates of the bacterial NANASs, LegSs and PseSs, contains an *N*-acetyl functionality at position C-2.

The absence of a catalytic Arg in the mammalian SAS family may indicate that a different mechanism is employed to position ManNAc-6-P for reaction with PEP. Further evidence that the interaction between a catalytic Arg and the *N*-acetyl group of ManNAc-6-P is not required in mammalian SASs, is given by the ability of *Hsa*NANA-9-PS to catalyse KDN-9-P synthesis by condensing mannose-6-phosphate with PEP (*182*). Conversely, none of the wild-type bacterial enzymes described in this thesis or in previous studies are reported to be capable of synthesising KDN from PEP and mannose, further implicating the necessity of the catalytic Arg interaction in NANAS, LegS and PseS enzymes. Of note, only the NANA-9-PSs of *H. sapiens* and *D. melanogaster* have been observed to exhibit KDN-9-PS activity (*30, 182*). Conversely, the NANA-9-PS of *Mus musculus* was shown to be unable to produce KDN-9-P (*187*).

Given the absence of the AFPL domain catalytic Arg in NANA-9-PSs, the residues which interact with the phosphate group of ManNAc-6-P may play a key role in both sugar substrate binding and reactive positioning. The K_m (ManNAc-6-P) for NANA-9-PS enzymes is significantly lower than that observed for the K_m (ManNAc) for NANAS enzymes (*103, 182*). It is possible that the role of the catalytic Arg is superseded by the residues which interact with the phosphate functionality of ManNAc-6-P in NANA-9PSs. Further experimental evidence would be required to compare the sugar substrate binding efficiencies of bacterial and mammalian SASs, such as ITC experiments using rManNAc and reduced ManNAc-6-P as well as site directed mutagenesis of key residues.

7.2.2 Pre-organisation of ManNAc binding site via a hydrogen bonding network

Not only does the AFPL domain donate a key active site residue in *Nme*NANAS, but it also plays an essential role in controlling ManNAc binding. The *Nme*NANAS AFPL domain truncation, *Nme*G272Term, was inactive and unable to bind ManNAc. Conversely, the truncation could still bind PEP, implying that the residues contributing to PEP binding are provided by the catalytic domain alone. Given that ManNAc binding is lost in *Nme*G272Term, but not in *Nme*R314A, the AFPL domain must be essential to binding ManNAc. As *Nme*Arg314 is non-essential to ManNAc binding, the AFPL domain likely contributes more than just a single essential residue into the active site.

The results obtained in Chapter 2 of this thesis reveal that *Nme*NANAS requires PEP to be bound before the enzyme can bind ManNAc, indicating that PEP liganding initiates a rearrangement event, pre-organising the ManNAc binding site. Interestingly, ITC results also reveal that PEP binding is entropically favourable – counter to the expected observation of an enthalpically favourable rearrangement. The entropically favoured rearrangement of an enzyme is not without precedent, with other examples of enthalpically unfavourable binding events previously reported (*188, 189*).

The largely favourable entropy in these examples is contributed by a hydrophobic effect where either water molecules are freed or hydrophobic surfaces are buried. Bezerra, G. A *et al* (*188*) describe an exceptionally large domain motion upon ligand binding in human dipeptidyl peptidase III, which is entropy driven. Dipeptidyl peptidase III releases water molecules from the substrate binding cleft of the enzyme upon substrate binding, a process which acts as a major thermodynamic driving force. The binding of PEP in *Nme*NANAS is accompanied by an observable conformational change as well as a significant thermal stabilisation of the enzyme. Given the structural analyses of the quaternary interfaces in the AFPL domain truncation and wild-type NmeNANAS enzyme, it is entirely plausible that the conformational change of *Nme*NANAS induced by PEP liganding involves a similar process which is driven by hydrophobic surface burial.

The absence of PEP induced conformational change and thermal stabilisation in *Nme*G272Term implicates the AFPL domains involvement in ManNAc binding site pre-organisation. Furthermore, MD simulations of PEP-bound wild-type *Nme*NANAS revealed that the flexibility of the ManNAc binding (β2-α2) loop and AFPL domain were restricted, enabling the catalytically critical ^{Nme}Arg314 residue to be held closer to the active site when compared to ligand-free wild-type. This observation further validates that conformational change upon PEP binding constructs the ManNAc binding pocket of *Nme*NANAS.

MD simulations and site directed mutagenesis analyses identified a hydrogen bond relay which links PEP binding to rearrangement of the AFPL domain. Residues ^{Nme}Glu282 and ^{Nme}Glu134 are each fully conserved in bacterial NANAS, LegS and PseS. The conservation of the hydrogen bonding network in bacterial SASs which bear AFPL domains, implies that the mechanism elucidated for ManNAc binding site pre-organisation in *Nme*NANAS is common to all members of the NANAS, LegS and PseS clades. Furthermore, *Cje*NANAS, *Cje*LegS, *Cje*PseS and *Ssu*SAS each undergo a similar conformational change and display a similar thermal shift in the presence of PEP when compared with wild-type *Nme*NANAS.

Discussed in Chapter 4, PEP binding did not appear to confer a significant conformational change in *Cte*SAS, which complies with the absence of an AFPL domain. Interestingly, even though *Cte*SAS does not contain an AFPL domain, PEP was observed to provide a significant amount of thermal stability to the enzyme. The lack of an AFPL domain in *Cte*SAS may indicate that members of the bacterial Type IV SAS clade have a unique mechanism for sugar substrate, binding site pre-organisation. Unfortunately, due to the unidentified sugar substrate and instability of *Cte*SAS, ITC experiments were problematic and it could not be determined as to whether or not PEP binding was required for the subsequent binding of a sugar substrate.

Whilst mammalian SASs do not lack an AFPL domain, the identified residues of the bacterial hydrogen bonding network are not conserved in NANA-9-PSs. Similar to each of the AFPL domain-bearing, bacterial SASs investigated in this thesis, *Hsa*NANA-9-PS was observed to undergo a conformational change and thermal shift in the presence of PEP. Given both the aforementioned observations and the presence of an AFPL domain in mammalian SASs, it is plausible that a similar network of residues is involved in a ManNAc-6-P binding site pre-organisation mechanism in the mammalian SAS family. Further investigation would be required to ascertain the exact residues involved, such as site-directed mutagenesis as well as molecular dynamics simulations of *Hsa*NANA-9-PS.

The existence of a similar PEP binding mechanism to *Nme*NANAS in *Hsa*NANA-9-PS is evidenced by the conformational change of the *Nme*/*Hsa* chimera in the presence of PEP. Since the catalytic domain is contributed by *Nme*NANAS and the linker and AFPL domain is contributed by *Hsa*NANA-9-PS, the typical interactions involved in either wild-type hydrogen bonding networks are not possible in the *Nme*/*Hsa* chimera. As a conformational change can still be induced by PEP binding, it is possible that the interactions between the *Nme*NANAS catalytic domain and *Hsa*NANA-9-PS linker and AFPL domain are similar enough to that of either wild-type species, so as to substitute for their normal hydrogen bonding relay interactions.

7.2.3 Stabilisation of quaternary structure

Lastly, the AFPL domain plays an important role in stabilising the quaternary state of *Nme*NANAS. The removal of the AFPL domain from *Nme*NANAS resulted in destabilisation of the protein's quaternary structure. The AFPL domain of *Nme*NANAS covers a hydrophobic cleft in the catalytic domain which is exposed in *Nme*G272Term. It is likely that the exposure of this hydrophobic core contributes to the protein's instability and eventual aggregation. Furthermore, the significant loss of intermolecular surface area contact in *Nme*G272Term when compared with wild-type *Nme*NANAS is responsible for destabilising the quaternary state of the truncated protein, which would in turn expose additional hydrophobic surfaces of *Nme*G272Term. The instability of *Nme*G272Term correlates with observations made of a previous truncation of the *Hsa*NANA-9-PS enzyme. The truncation of *Hsa*NANA-9-PS generated by Hao, J. *et al* (*103*), was also inactive and was noted to aggregate after a few hours. This indicates that the AFPL domain in *Hsa*NANA-9-PS plays an analogous role to that of the AFPL domain in *Nme*NANAS.

Similar to *Nme*G272Term, the *Nme*/*Hsa* chimera has a schizophrenic quaternary structure. Whilst the chimera is still able to retain some characteristics of the parent wild-type enzymes (such as conformational change induced by PEP binding), the *Nme*/*Hsa* chimera is not able to catalyse NANAS or NANA-9-PS reactions; even at concentrations where the chimera is predominantly dimeric. This is likely due to the incompatibility of the interfaces between the linker/AFPL domain and catalytic domain of the chimera. At high concentrations, the monomer-dimer equilibrium favours dimeric *Nme*/*Hsa* chimera, which may be able to overcome the energy barrier of the incompatibilities between the two interfaces. However, it is likely that the interfaces are too dissimilar to be able to complete a fully functional enzyme. The alternate, *Hsa*/*Nme* chimera is insoluble, which likely indicates that even fewer of the interactions are compatible between the *Hsa* catalytic domain and *Nme* linker/AFPL domain than between the *Nme* catalytic domain and *Hsa* linker/AFPL domain.

A previous chimera combining the linker and catalytic domain of *H. sapiens* NANA-9-PS with the AFPL domain of *M. musculus* NANA-9-PS (*Hsa*/*Mmu* chimera) was purified as a GST-fusion complex, and was determined to have both NANA-9-PS as well as KDN-9-PS activity (*187*). Whilst the *Hsa*/*Mmu* chimera had solubility issues it still retained activity, unlike the *Nme*/*Hsa* chimera. This is likely due to a much higher compatibility between the interfaces of the *Hsa*/*Mmu* chimera when compared with the *Nme*/*Hsa* chimera.

The two parent enzymes of the *Hsa*/*Mmu* chimera are more closely related with both source organisms being mammalian, whilst the source organisms of the *Nme*/*Hsa* chimera are more genetically distant. The *N. meningitidis* SAS shares only 28.5 % sequence identity with the *H. sapiens* SAS whilst the *M. musculus* SAS shares 94.7 % identity. The subunit interfaces in the *Hsa*/*Mmu* chimera would thus have a far greater chance of retaining interface contacts than that of the *Nme*/*Hsa* chimera.

As previously noted, members of the Type IV SASs naturally lack an AFPL domain. The results outlined in Chapter 4 of this thesis reveal that *Cte*SAS is a tight dimer even though the interdomain interactions are limited when compared with AFPL domain bearing SASs. It is likely that the interactions between the catalytic domains of the Type IV SASs have evolved to be much stronger than those of AFPL domain bearing SASs in order maintain a dimeric quaternary state. A crystal structure of *Cte*SAS would prove invaluable in determining the interactions of the dimer interface.

7.3 Evolutionary relationships of the sialic acid synthase family

7.3.1 The bacterial SAS family

Early investigation into the evolution of the SASs proposed that the enzymes were closely related to the PEP aldolase family (*122*). It was hypothesised that the SASs, DAH7PSs and KDO8PSs were descended from an ancestral PEP aldolase, as each of the previously mentioned enzymes perform very similar reaction chemistries. Furthermore, each of these PEP aldolases comprises a core $(\beta/\alpha)_8$ barrel catalytic domain. More recent investigations have dismissed this proposal, indicating that although the C-O bond cleavage mechanism is shared between these enzymes, it is only the DAH7PSs and KDO8PSs that are closely related (*113, 134*). The aforementioned investigations state that SASs appear to have a far more distant relationship to DAH7PSs and KDO8PSs, with the sequence identity of SASs to DAH7PSs and KDO8PSs being less than 10 % (*113*). Furthermore, the SASs have a distinct active site architecture as well as a unique barrel decoration that is not present in any DAH7PS or KDO8PS enzyme.

The work in this thesis reveals that the bacterial SAS family can be subdivided into four distinct clades: NANASs, LegSs, PseSs and Type IV SASs. As previously mentioned, NANAS, LegS and PseS bear AFPL domains and behave very similarly with regards to their PEP binding mechanisms. Whilst the AFPL domain is involved in the positioning of the sugar substrate in these enzymes it is likely that the residues provided by the β2-α2 loop provide substrate selectivity.

The Type IV SASs are unique in that they naturally lack AFPL domains. Originally, the Type IV SASs were proposed to be more closely related to the mammalian, NANA-9-PS enzymes and it was speculated that they synthesised NANA-9-P from PEP and ManNAc-6-P (*122*). The work in this thesis proved that this was not the case for *Cte*SAS, thus the true function of the Type IV SASs remains elusive.

An interesting observation made from the SAS phylogram described in Chapter 4, is the unusual placement of several SASs with known NANAS activity in the LegS clade. Both the SAS from *E .coli* and *S. agalacticae* have been previously observed to synthesise NANA (*95, 105, 158*). Furthermore, *Ssu*SAS (Chapter 5) has NANAS activity, yet is also placed in the LegS clade of the SAS family. *Ssu*SAS, *Sag*SAS and *Eco*SAS have high sequence identity to characterised LegSs such as *Lpn*LegS and *Cje*LegS. Unfortunately, LegS activity has not been tested for in *Ssu*SAS, *Sag*SAS or *Eco*SAS.

It is quite likely that LegSs are capable of catalysing both NANA and Leg biosynthesis. Described in both Chapter 4 of this thesis and a previous study (*64*), *Cje*LegS is observed to have minimal NANAS activity. In addition, complementation studies have indicated that *Lpn*LegS could restore the biosynthesis of NANA in an *E. coli* gene knockout mutant, suggesting that the SAS from *L. pneumophila* could synthesise either Leg or NANA (*136*). In addition, modelling of *Ssu*SAS indicates that the enzyme's active site is better suited to accommodate 2,4-diacetamido-2,4,6-trideoxymannose than ManNAc.

The ability of LegSs to catalyse the biosynthesis of both NANA and Leg, may provide insight into the ancestry of the SAS family. It is possible that NANAS enzymes such as those from *N. meningitidis* and *C. jejuni* evolved from LegSs which are capable of catalysing both NANAS and LegS reactions. The evolutionary pressure to mask pathogenic organisms such as *N. meningitidis* and *C. jejuni* from host immune responses may have driven the development of NANAS specific enzymes. This hypothesis would need to be further examined, as LegS and PseS activity have not been examined for either *Nme*NANAS or *Cje*NANAS.

Similarly, it is possible that the PseS clade evolved from an ancestral LegS, as PseSs are more closely related to LegSs than NANASs or Type IV SASs. The substrate of LegS is quite similar to that of PseS, but whether or not LegSs can catalyse the formation of Pse, or PseSs the formation of Leg, is currently unknown. Described in Chapter 4, *Cje*PseS cannot catalyse the formation of NANA. Both Leg and Pse are involved in the motility of prokaryotes, which is arguably a more fundamental and less niche evolutionary driving force than host immune evasion. It is thus possible that the enzymes required for the synthesis of Leg and Pse may have evolved before those involved in NANA biosynthesis.

The evolution of Type IV SASs is particularly interesting, as the proteins of the Type IV clade are distinct from the other members of the SAS family. The absence of the AFPL domain from the Type IV SASs may indicate that the decoration was lost due to its lack of necessity to the Type IV catalytic mechanism, or that Type IV SASs resemble an ancestral enzyme, prior to the acquisition of the AFPL domain in SAS evolution. The latter theory is supported by the previous findings of Lewis, A. L. *et al* (*134*) who identify several archael sequences in the Type IV clade and similarly suggest that the members of the Type IV clade are of a more ancient origin than the LegS, PseS and NANAS enzymes.

The preceding arguments for the evolution of the bacterial SASs ignore the possibility of paralogous gene transfer between mammalian and bacterial SASs. Due to the similarity of NANA biosynthesis in both bacteria and mammals, it was once believed that the genes for bacterial NANASs had been acquired from the mammalian paralogue through horizontal gene transfer or vice versa (*122*). This is exceedingly unlikely given the wide distribution of bacterial and mammalian SASs, as well as the presence of several archael sequences in each of the SAS clades from the phylogram described in Lewis, A. L. *et al* (*134*). The presence of archael sequences in each of the SAS clades indicates their divergence occurred long ago. Furthermore, there is a significant lack of sequence similarity between NANA and NANA-9-PS enzymes. Bacterial NANASs are far more closely related to other bacterial SASs than to mammalian NANA-9-PSs. As previously mentioned, NANA-9-P biosynthesis involves phosphorylated metabolites and an additional enzyme to phosphorylate and dephosphorylate ManNAc. Conversely, none of the bacterial sialic acids are known to utilise similar routes for biosynthesis.

Lewis, A. L. *et al* (*134*) propose an alternative theory for the evolution of both bacterial and mammalian SASs, stating that present day SASs, which initially appeared to have an animal heritage, are more likely to be the present-day counterparts of an ancient pathway that was universally adopted by deuterostome animals. Lewis, A. L. *et al* also suggest that SAS biosynthesis in animals did not arise by lateral gene transfer, but was instead inherited in the traditional sense, accompanied by multiple gene losses along other eukaryotic lineages.

7.3.2 Evolution of the antifreeze protein fold

The theory of SAS evolution proposed by Lewis, A. L. *et al* (*134*) is particularly interesting with respect to the evolution of the antifreeze protein fold which is present in both bacterial and mammalian SASs as well as Type III antifreeze proteins (*115, 117*). Based upon this theory, it is likely that the antifreeze protein fold originated in the SAS family and was eventually commandeered and modified to resemble modern, Type III antifreeze proteins.

This correlates with the propositions of Baardsnes, J. *et al* (*117*) who observed that not only is the SAS AFPL domain related to Type III fish antifreeze proteins, but that many other types of antifreeze proteins have been evolutionarily repurposed. Type II antifreeze proteins have been thought to have evolved from lectins, plant antifreeze proteins have been thought to have evolved from enzymes involved in polysaccharide hydrolysis, and carrot antifreeze proteins are evolutionarily related to a polygalacturonase inhibitor (*117*). Several modern antifreeze proteins are thus modified from proteins involved in sugar binding, which correlates with the theory that Type III antifreeze proteins may have been similarly adapted from the SAS AFPL domain. Baardsnes, J. *et al* also speculated that the hydroxyl groups of sugars and polysaccharides resemble the oxygen atoms on certain ice planes, and predicted that the AFPL domain of the SAS family was involved in sugar binding, prior to the elucidation of the *Nme*NANAS structure 1XUZ, and long before the findings reported in this thesis.

7.4 Future work

7.4.1 Drug Design

As previously mentioned, NANAS is essential to the pathogenesis of *N. meningitidis*, *H. influenza*, *E. coli* K1 and *C. jejuni*. Given the essential role of the catalytic Arg residue in bacterial SASs and its absence in the mammalian NANA-9-PS enzyme, inhibitors could be designed to take advantage of this difference by binding and removing access to the Arg guanidinium functionality. A previous analysis of *Sag*SAS revealed that activity could be inhibited with the addition of phenylglyoxal, a compound which covalently modifies accessible arginine residues (*105*). Inhibitors could also be derived from PEP analogues, as has been previously described for both DAH7PS and KDO8PS enzymes (*190, 191*), and engineered to be specific to bacterial SASs. Currently, the only inhibitor designed for *Nme*NANAS is the tetrahedral intermediate mimic described by Liu, F. *et al* (*113*).

7.4.2 Crystallisation

Currently, the structures of *Nme*NANAS and *Ssu*SAS are the only complete SAS crystal structures known, and only the AFPL domain of *Hsa*NANA-9-PS has been elucidated. The determination of additional SAS structures will be essential to uncovering evolutionary differences and identifying the residues involved in substrate selectivity between these enzymes. While it is yet to be determined whether or not *Ssu*SAS is a LegS, there is no structure of a PseS, Type IV SAS or NANA-9-PS currently available. The work in this thesis has extensively trialled a range of crystallisation conditions for several enzymes, however, further work may still be done in this area. In particular, the thermal stability of SASs in the presence of Mn^{2+} and PEP may provide assistance in crystallisation. Although the enzymes described in this thesis were trialled with ligand present, conditions could still be optimised to provide diffraction quality crystals.

7.4.3 Evolution of the SAS family

The work in this thesis outlines a preliminary investigation into the evolutionary history of the SAS family. Further analyses are required to probe the relationships within the SAS family and determine the ancestry of each clade. The work outlined in this thesis proposes that LegSs may be capable of catalysing both NANAS and LegS reactions. Several bacterial SASs, identified as LegSs in this thesis, have been denoted as NANASs in online databases, however, their ability to synthesise Leg has never been determined. Similarly, whether or not NANASs can catalyse LegS or PseS reactions and whether PseSs can catalyse LegS reactions or vice versa, has never been fully investigated. The absence of information in this regard is likely a result of the sugar substrates for LegSs (2,4-diacetamido-2,4,6-trideoxymannose) and PseSs (2,4-diacetamido-2,4,6-trideoxaltrose) not being commercially available and requiring several steps to be chemically synthesised (*161*). The work outlined in Chapter 5 of this thesis strongly indicates that *Ssu*SAS is capable of LegS activity, however, kinetic analysis with 2,4-diacetamido-2,4,6-trideoxymannose would solidify these claims.

The most direct method to determine whether or not different members of the SAS family are able to synthesise a variety of reactions, is through kinetic characterisation. However, this could also be proven through gene complementation studies. The restoration of a NANA biosynthetic pathway in a gene knockout mutant has been previously examined with *Lpn*LegS (*136*), and similar experiments with other SASs could be used to the same effect. The determination of substrate plasticity in the SAS family would provide valuable insight into the evolution of the different clades.

7.4.4 Can SAS activity be interconverted?

Another interesting avenue in investigating the evolutionary relationships of the SAS family, is to determine which specific residues define the members of each clade and control substrate selection. Outlined in Chapter 4 of this thesis, there are two specific areas where the peptide sequence varies between clades; the first is a section on the AFPL domain linker and the second is the β2-α2 loop. Both regions provide residues proximal to the sugar binding site and are likely involved in sugar substrate recognition. Residues identified from the models and substrate docking experiments of *Cje*NANAS, *Cje*LegS and *Cje*PseS in Chapter 4 may provide a basis for site-directed mutagenesis experiments to interconvert the activities of these enzymes. Similarly, chimeric enzymes could be generated which interchange the AFPL domain linker regions or β2-α2 loops.

The work outlining the *H. sapiens* and *N. meningitidis* chimeras in Chapter 6 identifies issues involved with interchanging the AFPL domains of different SASs. The compatibility of interdomain contacts needs to be examined thoroughly when constructing SAS AFPL domain chimeras. It is possible that the stability of SAS chimeras could be improved through site directed mutagenesis along the dimer interface, in an effort to strengthen interdomain interactions.

The lack of NANAS activity in *Cte*SAS is likely a result of the absence of an AFPL domain. It may be possible to confer NANAS, LegS, PseS or NANA-9-PS activity by incorporating an AFPL domain from a member of a selected SAS clade. As aforementioned, the naturally truncated form of *Cte*SAS may be indicative of an ancestral SAS enzyme, and the ability of *Cte*SAS to incorporate an AFPL domain may provide some insight into the evolution of the unique antifreeze protein fold.

7.4.5 What is the reaction catalysed by Type IV SASs?

As previously mentioned, the sugar substrate of the Type IV SAS clade has yet to be determined. Further kinetic analysis of alternative sugars with *Cte*SAS or another Type IV SAS would be needed to identify the compound that is reacted with PEP. ITC could be used to determine whether particular sugar molecules are able to bind to *Cte*SAS and provide a profile to narrow down possible targets. The crystallisation of a Type IV SAS would provide invaluable insight into the catalytic site and sugar binding pocket. Molecular modelling could then be used to investigate or probe the sugar binding site, and elucidate structural features pertaining to the substrate molecule. Another method would be to determine which sialic acid compounds are present in a selected organism which expresses a Type IV SAS, using NMR and mass spectroscopy. By identifying the sialic acids which are present in the organism, the identity of the sugar substrate that is reacted with PEP could be retrosynthetically determined.

7.5 Concluding remarks

Sialic acids are almost ubiquitous in nature and are involved in a wide variety of intercellular processes. Sialic acids play a key role in the pathogenesis of neuroinvasive pathogens and as such, bacterial SASs such as NANAS are exemplary drug targets. Little is currently understood about the evolution of SASs which are widespread; present in both bacteria and mammalian species. The work outlined in this thesis has afforded considerable insight into the catalytic mechanism of *Nme*NANAS, and the role of the structurally unique, AFPL domain. Understanding the mechanism of SAS enzymes provides a valuable model for drug design against pathogenic organisms such as *N. meningitidis* and *C. jejuni*. Furthermore, a second SAS crystal structure has been elucidated, and an investigation into the evolutionary relationships of the SAS family has uncovered four distinct SAS clades and identified structural variances, which may control substrate selectivity.

Chapter 8

Materials and Methods

8.1 Figures and alignments

8.1.1 Protein structure figures

Figures of protein structures were created using PyMOL (version 1.5, Schrödinger LLC) (*192*).

8.1.2 Multiple sequence alignments

Multiple sequence alignments were generated using Tcoffee (*193*) or ClustalW (*155*) and formatted for display using Jalview (*159*).

8.1.3 Generation of phylograms

ClustalW alignments were used to generate a distance matrix using the default settings. The distance matrix was subsequently entered into Figtree (*157*) to generate the phylograms displayed.

8.2 Reagent preparation

8.2.1 Water

All water used in these experiments was from a Millipore Milli-Q system. This water is referred to as Milli-Q water. Water used for PCR and crystallisation experiments was autoclaved prior to use.

8.2.2 pH Measurements

pH measurements were made using a Denver Instruments UB-10 Ultra-Basic pH meter, with a standard probe. Solutions were acidified by addition of 1 M or 10 M HCl and alkalised by addition of either 1 M or 10 M NaOH. When chloride ions were to be avoided (for circular dichroism experiments) H_2SO_4 was used instead of HCl. Unless otherwise noted, buffers were pH adjusted at room temperature.

8.2.3 Removal of metal ions from solutions

Divalent metal ions were removed from buffers used for metal dependency experiments by treatment with Chelex® 100 resin (Bio-Rad). The resin was added to the buffer solution at a ratio approximately 5 g per 100 mL of buffer and then stirred for approximately 1 hour. Resin was subsequently filtered under vacuum (0.2 µm filter). The resin invariably caused the pH of the solution to increase, often necessitating adjustment to correct the pH of the Chelex treated solution. Glassware used during and after chelex treatment were washed with concentrated nitric acid.

8.2.4 Determination of substrate concentrations

Concentrations of PEP and ManNAc solutions were determined using the 232 nm assay with *Nme*NANAS. Known volumes of a solution containing the substrate for which the concentration was to be measured were added to the assay cuvette while the second substrate was kept at excess concentration. To confirm the desired substrate was limiting, after the enzyme catalysed reaction was complete more was added to the assay.

The observed change in absorbance was measured in duplicate as ΔA_{rxn}. A control sample lacking substrates was used to measure the change in absorbance upon enzyme addition (ΔA_{enz}). The corrected final change in absorbance, calculated as $\Delta A_{rxn} - \Delta A_{enz}$, was converted to the concentration of the limiting substrate by the Beer-Lambert law, using the extinction coefficient of PEP (2.8×10^3 $M^{-1} cm^{-1}$).

8.2.5 Antibiotics

Ampicillin was added to all growth media to a final concentration of 0.1 mg/mL. Stock solutions of ampicillin at 100 mg/mL dissolved in Milli-Q were created, filter sterilised and stored at -80 °C, and freeze-thawed a maximum of two times. The *E. coli* Rosetta™ 2 (DE3) expression cell line contains resistance for chloramphenicol and the chaperone 3 expression cell line contains resistance to chloramphenicol and spectinomycin. When these cell lines were used, appropriate antibiotics were added to all growth media in addition to ampicillin, to a final concentration of 25 µg/mL for chloramphenicol and 0.1 mg/mL for spectinomycin. Stock solutions of chloramphenicol were dissolved in ethanol at 25 mg/mL and spectinomycin stocks were dissolved in Milli-Q at 10 mg/mL. Antibiotics were filter sterilised and stored at -80 °C, and freeze-thawed a maximum of two times.

8.2.6 Culture media

Lysogeny-broth (LB) agar was prepared by dissolving LB (Lennox L) (20 g L^{-1}) and agar (15 g L^{-1}) base in Milli-Q water and sterilised by autoclaving. The LB-agar solution was then heated in a microwave oven until boiling, and left to cool in a 60 °C oven. The LB-agar solution was then further cooled to 50 °C before antibiotic(s) were added immediately prior to pouring into round petri dishes.

LB media for plasmid extraction, pre-cultures and protein expression cultures were prepared by dissolving 20 g L^{-1} LB (Lennox L) base in Milli-Q water and sterilised by autoclaving. Antibiotic(s) were added immediately prior to use.

8.3 Cloning and transformation

8.3.1 Primers

Primers were designed by hand or by PrimerX (www.bioinformatics.org/primerx) and synthesised by GeneWorks or Invitrogen. Primers were stored at 300 mM in autoclaved TE-buffer (10mM Tris-HCl, 0.1mM EDTA).

8.3.2 PCR

Polymerase chain reactions (PCR) were performed using a Veriti®96-well Thermal Cycler (Applied Biosystems) or an iCycler (Bio-Rad). PCR was performed using *PfuUltra* High-Fidelity DNA Polymerase (Agilent). A 50 µL mixture was created for each reaction using the combination outlined in Table 8.1. The amplification protocol used for PCR is outlined in Table 8.2. PCR products were purified using E-Gel® CloneWell™ gels (Life Technologies).

Component	Amount per reaction
Distilled water (dH_2O)	40.6 µL
PfuUltra Reaction buffer	5.0 µL
dNTP mix (25 mM each dNTP)	0.4 µL
DNA template (100 ng/µL)	1.0 µL
Fwd primer (100 ng/µL)	1.0 µL
Rev primer (100 ng/µL)	1.0 µL
PfuUltra DNA Polymerase (2.5 U/µL)	1.0 µL
Total reaction volume	50.0 µL

Table 8.1 PCR reaction mixture for *PfuUltra* High-Fidelity DNA Polymerase (Agilent).

Number of Cycles	Temperature	Duration
1 x	95 °C	1 minute
30 x	95 °C	30 seconds
	40-60 °C	30 seconds
	72 °C	1 minute per kb
1 x	72 °C	10 minutes

Table 8.2 PCR amplification protocol for *PfuUltra* High-Fidelity DNA Polymerase (Agilent).

8.3.3 Genomic cloning

The gene encoding *Nme*NANAS was amplified from *N. meningitidis* MC58 (serogroup B) genomic DNA (ATCC) whilst genes encoding *Cje*NANAS, *Cje*LegS and *Cje*PseS were amplified from *C. jejuni* NCTC 11168 genomic DNA (ATCC).

Cycling parameters and components used for the 50 µL reaction are outlined in Section 8.3.2, but with the introduction of a touchdown annealing stage, wherein the annealing temperature of the PCR cycle started at 70 °C and dropped 1 °C each cycle for twenty cycles. The annealing temperature for the remaining 10 cycles was 40 °C. Resulting PCR fragments were assessed for appropriate size by agarose gel electrophoresis and ligated into vector pDONR-221 as described in Section 8.3.5.

8.3.4 Mutagenesis

Mutagenesis was performed using either a Quikchange® Lightning Mutagenesis Kit (Stratagene), Quikchange® XL Mutagenesis Kit, or with *PfuUltra* High-Fidelity DNA Polymerase (Agilent). PCRs were performed as outlined in Section 8.3.2. Template DNA (which is methylated) was digested by treatment with the restriction enzyme DpnI (Life Technologies).

8.3.5 Agarose gel electrophoresis

Agarose gel electrophoresis was performed using either pre-cast E-Gel® 1.2 % (w/v) agarose gels (Invitrogen) or self-poured 1 % agarose gels. E-Gel® agarose gels were used as per manufacturer's instructions. Samples were added directly to the gel, without loading dye. Gels were run on and monitored with an E-Gel R iBase™ Safe Imager™.

Self-poured gels were made by heating 0.3 g agarose in 30 mL tris-acetate-EDTA (TAE) buffer until dissolved. The solution was cooled to approximately 50 °C before being supplemented with SYBR Safe® DNA stain and poured. For self-poured gels, samples were mixed with a 6 x sample loading dye. Electrophoresis was performed for 60 min at 80 V in TAE buffer using a Mini-Sub® Cell GT (Bio-Rad). All gels were viewed and photographed under UV light with a Molecular Imager® Gel Doc™ XR (Bio-Rad).

TAE buffer: 60mM Tris-HCl, 1mM EDTA, 20mM acetic acid

Sample loading dye (6 x): 60mM Tris-HCl, 60mM EDTA, 0.2% (w/v) orange G, 0.05% (w/v) xylene cyanol ff, 60% (v/v) glycerol

8.3.6 Cell lines

The cell lines used for plasmid propagation were *E. coli* TOP10 cells (Invitrogen). For protein expression *E.coli* BL21 (DE3) cells, *E. coli* Rosetta™ 2 (DE3), or *E. coli* BL21 (DE3) cells containing the plasmids pBB528 and pBB541 for co-expression of the chaperonins GroEL and GroES were used.

8.3.7 Preparation of chemically competent cells

Chemically competent cells were prepared by two similar methods. In the first protocol, cells were prepared using the method of Hanahan (*194*). An overnight culture of the cells to be made competent was grown in the presence of the appropriate antibiotic(s). The following day lysogeny broth (LB) medium with the appropriate antibiotic was inoculated with one hundredth the volume of the overnight culture. Cells were grown at 37 °C until reaching an OD_{600} of 0.3 AU to 0.6 AU. The culture medium was then incubated on ice for 10 - 15 min, before being harvested by centrifugation for 15 min at 4 000 rpm and 4 °C, in a pre-chilled centrifuge and rotor. The pellet was then resuspended in 18 mL of RF1 solution and incubated on ice for 30 min. The cells were then harvested again by the same centrifugation method, and resuspended in 4 mL of RF2 solution. Suspended cells were then split into 50 µL to 100 µL aliquots, flash-frozen in liquid nitrogen and stored at -80 °C.

RF1 solution: 100 mM $RbCl_2$, 50 mM $MnCl_2$, 30 mM KOAc, 10 mM $CaCl_2$, pH 5.8 (adjusted using concentrated acetic acid).

RF2 solution: 10 mM $RbCl_2$, 10 mM MOPS, 75 mM $CaCl_2$ · 6H2O, 15% (v/v) glycerol, pH 5.8 (adjusted using NaOH).

The RF1 and RF2 solutions were filter-sterilised (0.2 µm) rather than autoclaved to avoid formation of a Rb precipitate, and stored at 4 °C.

In the second method to prepare competent cells, an overnight culture was grown including the appropriate antibiotic(s). The next morning, a 100mL culture was inoculated with 2mL of the overnight culture, and grown at 37 °C until OD_{600} reached 0.3 - 0.4 AU. At this point the cells were transferred to a sterile 50 mL centrifuge tube, and centrifuged at 4 000 rpm for 15 min at 4 °C to pellet the cells. The supernatant liquid was poured off and the pellet gently resuspended on ice in 50 mL of freshly-made sterile 100 mM $CaCl_2$. The resuspended cells were left on ice for 30 min, after which they were again centrifuged, and the pellet gently resuspended in 10 mL of cold $CaCl_2$ on ice. The cells were then finally incubated on ice for at least 2 h, before being gently mixed with pre-chilled sterile 60% (v/v) glycerol to a final concentration of 15% (v/v). The cells were then split into 50 µL to 100 µL aliquots, snap-frozen in liquid nitrogen, and stored at - 80°C.

8.3.8 BP clonase reaction

A BP reaction was used to transfer linear PCR product into the donor vector pDONR221 for subsequent use of the Gateway cloning system (Life Technologies). BP Clonase (Invitrogen) was used according to the manufacturer's instructions, and the reaction product was transformed into TOP 10 cells for plasmid proliferation using the method outlined in Section 8.3.10. The cells were grown on agar plates containing 50 μg/mL kanomycin and the plasmids extracted for sequence verification followed by subcloning into expression vectors using the LR reaction outlined in Section 8.3.9.

8.3.9 LR clonase reaction

An LR reaction was used to subclone the target gene into a selected destination vector via pDONR221. Destination vectors used were pDEST15 which contains an N-terminal GST-tag sequence, pDEST17 which contains an N-terminal His-tag sequence and pDEST566 which contains an N-terminal MBP-tag sequence. LR Clonase (Invitrogen) was used according to the manufacturer's instructions, and the reaction product was transformed into TOP 10 cells for plasmid proliferation using the method outlined in Section 8.3.10. The cells were grown on agar plates containing 100 μg/mL ampicillin and the plasmids extracted for subsequent transformation into expression cell lines for purification.

8.3.10 Transformation

Chemically competent cells (100 μL to 200 μL aliquots) were thawed on ice before addition of 2 μL of either PCR product or purified plasmid. The mixture was left on ice for 30 min then placed in a 42 °C water bath for 30 s to 45 s before being placed on ice again for 2 min. Cells were then outgrown for 1 hr at 37 °C while shaking after addition of 250 μL super optimal broth (SOC) or LB medium. A portion (50 μL to 200 μL) of the transformed cells was spread directly onto an LB-agar plate containing the appropriate antibiotic(s) and left to grow on the inverted medium overnight at 37 °C.

SOC medium: 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 20 mM glucose. Filter sterilised in aliquots.

8.3.11 Plasmid extraction and purification

Plasmids were extracted using a High Pure Plasmid Isolation Kit (Roche). Small cultures (5 mL) of plasmid-containing cells were grown overnight with the appropriate antibiotic(s). The cultures were harvested the following morning and plasmids extracted and purified as per the kit instructions. The concentration of purified plasmid was measured by absorption at 260 nm using an appropriately blanked Nanodrop ND-1000 spectrophotometer.

8.3.12 Colony PCR

After transformation of E. coli TOP10 cells with pDONR221plasmids, promising candidates for sequencing were identifed with colony PCR. This procedure used the forward gene specific DNA sequencing primer and reverse T7 sequencing primer (Table 8.3) to confirm that the subcloned gene had been inserted into the vector in the correct orientation. This was confirmed by the amplification of an appropriately sized DNA fragment from the colony PCR.

PCR reactions were set up using Taq DNA Polymerase (Invitrogen) and following the manufacturer's instructions for reaction components and cycling parameters. Template DNA was provided as a toothpick scraping of an E. coli TOP10 colony from an agar plate of suitably-transformed cells.

Oligonucleotide	Sequence (5' - 3')
T7 Fwd	TAATACGACTCACTATAGGG
T7 Rev	GCTAGTTATTGCTCAGCGG

Table 8.3 T7 primer sequences.

8.3.13 DNA sequencing

DNA sequencing was used to verify amplified gene sequences after subcloning into pDONR221. Sequencing was performed by Canterbury Sequencing on an ABI3100 Genetic Analyzer (Applied Systems Inc.) using a procedure based on Sanger chain termination protocol. Typically, double-stranded plasmid samples were prepared as 6 mL aliquots of 40 ng/mL plasmid per sequencing run. T7 Primers were used for sequencing at a concentration of 3.2 mM (Table 8.3). aliquots.

8.3.14 Glycerol stocks

In order to keep cells viable, all strains of *E. coli* cells containing plasmids were flash frozen and stored at -80 °C as glycerol stocks. Glycerol stocks were created by adding 400 µL of 50% (v/v) glycerol with 800 µL of an overnight culture in either a 1.6 mL eppendorf or screw-top tube.

8.4 Expression and purification

8.4.1 IPTG-induced expression

1L LB media cultures for protein expression were grown in 2 L baffled conical flasks in a shaking incubator at 170-180 rpm. Cultures were inoculated with an overnight culture 1/20 of the expression volume and grown at 37 °C until induction of protein expression. The OD_{600} of the cultures was monitored until reaching mid-logarithmic phase (OD_{600} of 0.4-0.8 AU), which generally took 2-4 hours. At this point isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. After induction, cultures were transferred to a 23 °C incubator shaking at 170-180 rpm and grown overnight.

Cultures used for expression trials were grown using 50 mL of LB media in 300 mL conical flasks.

8.4.2 Cell harvesting

The method used to harvest cells from cultures used for protein expression varied depending on rotor and centrifuge availability. Typically, large cultures were harvested in 0.8 L or 1 L bottles at 14 000 *g* for 30 min at 4 °C. The supernatant liquid was removed and pellets were then transferred to 50 mL centrifuge tubes and stored at -80 °C. Alternatively, cultures were sometimes progressively harvested in 50 mL centrifuge tubes at 12 000 *g* for 10 min at 4 °C, or in 0.4 L buckets at 4 618 *g* for 40 min at 4 °C before being transferred to 50 mL centrifuge tubes and stored at -80 °C. Small volume cultures (5 mL) were progressively harvested in 1.6 mL micro-centrifuge tubes at 17 000 *g* for 3 min.

8.4.3 Cell Lysis

Cells were lysed by sonication using an Omni-Ruptor 4000 Ultrasonic Homogenizer (Omni International). Cell pellets were resuspended in chilled lysis buffer (typically 50 mL to 100 mL) on ice, and sonicated in a beaker surrounded by packed ice. Sonication was performed typically for 20 min at 40% pulsation and 70% power. Sample temperature was checked every 5 min to prevent over-heating. Lysis buffer was varied from purification to purification and is addressed in Section 8.4.4.

For small volumes such as with expression tests, Bugbuster® protein extraction reagent (Novagen) was used in place of sonication. Cell pellets were resuspended in a smaller volume of buffer (approximately 5 mL) containing Bugbuster® reagent. 5 µL of Benzonase nuclease® (Novagen) was added to break down DNA, and the solution shaken at 25 °C for 20-30 mins.

After lysis, the soluble fraction was separated from the cell debris by centrifugation at 24,000 g for 30 min at 4 °C. Chromatography was performed as quickly as possible after lysis to reduce the effects of proteolysis.

8.4.4 Chromatography

All proteins were purified by a multi-step procedure outlined in Figure 8.1. The chromatography steps were performed on Bio-Rad Biologic DuoFlow and GE Healthcare ÄKTApurifier™ 10 machines.

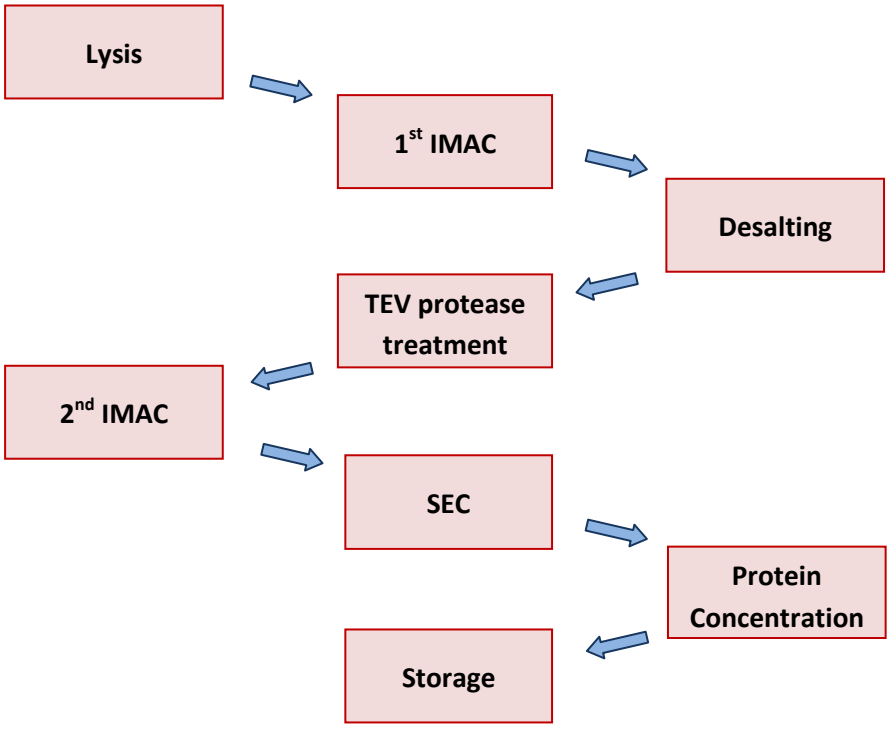


Figure 8.1 Purification procedure overview.

Before use, all buffers were filtered (0.2 µm) under vacuum. Protein samples were similarly filtered before being loaded on either a 50 mL or 200 mL Superloop™ (GE Healthcare). The eluate was fractionated in 2 mL fractions collected in 96-well plates. Elution from the columns was monitored principally at 280 nm, and depending on the machine being used, also at 260nm and 214 nm. When the fractions corresponding to the target protein could not be deduced from the chromatogram, suspected fractions were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and/or for NANAS activity when possible. Fractions containing target protein were subsequently pooled and prepared for the next chromatographic step.

Given the variety of proteins expressed and purified in this thesis, the same buffers could not always be used for purification. For the majority of wild-type and variant proteins the following buffers were used:

Lysis buffer: 50 mM potassium phosphate (pH 7.5), 500 mM NaCl, 10 mM imidazole

Talon equilibration: 50 mM potassium phosphate (pH 7.5), 500 mM NaCl, 10 mM imidazole

Talon elution buffer: 50 mM potassium phosphate (pH 7.5), 500 mM NaCl, 200 mM imidazole

Size exclusion buffer: 30 mM triethanolamine (pH 7.5)

*Cte*SAS was found to be unstable in the absence of PEP, thus 0.1 mM PEP was added to the buffer solutions for the purification of *Cte*SAS. In addition, *Cte*SAS was unstable in triethanolamine and was instead purified by SEC in 20 mM BTP and 150 mM NaCl at pH 7.5.

8.4.5 Immobilised metal affinity chromatography

Purification of all proteins described in this thesis involved two IMAC steps (Figure 8.1). The first of these separated the His-tagged protein from the soluble *E. coli* proteins, and the second separated untagged protein from the tag and TEV protease after cleavage. IMAC was performed using 5mL Talon® Superflow Metal Affinity resin packed in a Tricorn 10/50 column (GE Healthcare). The resin was equilibrated with five column volumes of equilibration buffer prior to injection of protein sample.

Resin was washed with ten column volumes of equilibration buffer, before elution of bound protein with six column volumes of elution buffer. Following elution, the resin was washed with five column volumes of 2-(N-morpholino)ethanesulfonic acid (MES) buffer (pH 5.5), containing 200 mM NaCl, to remove bound imidazole before being stored in ethanol. The protein of interest was collected from either the eluted fraction (for the first IMAC step) or the washing step (for the second IMAC step) depending on the presence of the protein His-tag. The elution flow rate used for IMAC was 3 mL min^{-1}

8.4.6 Desalting

Desalting was performed using a HiPrep 26/10 Desalting column pre-packed with Sephadex G-25 Fine (GE Healthcare). Desalting chromatography was carried out at 4 °C in low imidazole buffer (50 mM potassium phosphate at pH 7.5, 500 mM NaCl, 10 mM imidazole) to remove high imidazole concentrations subsequent to TEV protease treatment and also prior to the second IMAC run. The elution flow rate used for desalting was 2 - 5 mL min^{-1}.

8.4.7 TEV protease treatment

Each of the proteins mentioned in this thesis contained cleavage sites for tobacco etch virus (TEV) protease, which was used to remove affinity tags. The cleavage sequence used was ENLYFQ/G, with TEV protease cleaving between the glutamine and glycine residues. To cleave the His-tag from the protein of interest, affinity-purified and desalted protein was incubated at 4 °C overnight with TEV protease. This reaction was performed in Talon equilibration buffer (Section 8.4.4) with the addition of 1 mM DTT and 0.5 mM EDTA. TEV protease was added at a ratio of 1:100 to the substrate protein (by mass, as determined by absorbance).

8.4.8 TEV protease purification

Recombinant TEV protease was purified similarly to the purification protocol outlined in Section 8.4.4. *E. coli* BL21(DE3) cells containing the plasmids pRIL (for rare codon production) and pRK793 (encoding TEV protease) were grown in LB containing ampicillin and chloramphenicol. Protein over-expression was induced with 0.5 mM IPTG and the cells grown overnight. Cells were harvested and lysed as outlined in Section 8.4.3, using the lysis buffer described below. Protein was then purified in a single IMAC step on a HisTrap HP column (GE Healthcare). IMAC procedures were identical to those used for Talon® resin (Section 8.4.5), using the buffers described below. Purified TEV protease was buffer exchanged into storage buffer, concentrated to 1 mg/mL and stored in 1 mL aliquots.

TEV protease lysis/equilibration buffer: 50 mM potassium phosphate, pH 8.0, 500mM NaCl, 25 mM imidazole

TEV protease elution buffer: 50 mM potassium phosphate (pH 8.0), 500 mM NaCl, 250 mM, 500 mM imidazole

TEV protease storage buffer: 10% (v/v) glycerol, 25 mM potassium phosphate (pH 8.0), 200 mM NaCl, 10 mM DTT, 2 mM EDTA

8.4.9 Size exclusion chromatography

Pooled fractions containing target protein from the second IMAC run were pooled and loaded onto a HiLoad™ 26/60 Superdex™ 200 prep grade column (GE Healthcare). SEC chromatography was carried out at 4 °C in 30 mM triethanolamine buffer. The elution flow rate used for SEC purification was 0.5 - 1 mL min^{-1}.

8.4.10 SDS-PAGE electrophoresis

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using pre-cast gels using NuPAGE® 10% Bis-Tris Gel 1.0mm × 12-well pre-cast gels (Invitrogen) in NuPAGE® MOPS SDS Running Buffer (Invitrogen). Electrophoresis was performed using a XCell SureLock™ Electrophoresis Cell (Invitrogen), at 200 V for 50 min. Samples were mixed with NuPAGE® LDS Sample Buffer (4×) and DTT, and boiled before loading.

8.4.11 SDS-PAGE visualisation

All SDS-PAGE gels were stained in square petri dishes by coomassie brilliant blue R-250, for at least 1 hour before being destained to remove excess dye.

Stain: 0.1% (w/v) coomassie brilliant blue R-250, 10% (v/v) acetic acid and 40% (v/v) methanol

Destain: 10% (v/v) acetic acid and 40% (v/v) methanol

Molecular weight standards (Novex® Sharp Pre-Stained Protein Standards (Invitrogen)) were run as the sample in one lane of each gel. Photographs of gels were taken on a white light transilluminator in a Molecular Imager® Gel Doc™ XR (Bio-Rad).

8.4.12 Protein concentration and buffer exchange

Protein solutions were concentrated using 10 000 Da molecular weight cut-off (MWCO) devices (Vivaspin 2 (GE Healthcare), Vivaspin 500 and Vivaspin 20 (Sartorius Stedim Biotech) and Amicon® Ultra-4 (Millipore)). All filtration units were rinsed before use with Milli-Q water. Protein solutions were buffer exchanged by repeatedly concentrating in a MWCO device and diluting with desired buffer.

8.4.13 Enzyme storage

Purified protein was divided into aliquots in either 0.6 mL or 1.2 mL micro-centrifuge tubes in 0.1 mL to 0.5 mL aliquots, flash-frozen in liquid nitrogen, and stored at -80 °C. All protein samples were rapidly thawed prior to use and kept on ice. Protein solutions were stored at 1.0 mg mL^{-1} to 15 mg mL^{-1}.

8.4.14 Protein concentration determination

Protein concentrations were measured by absorption at 280 nm by an appropriately blanked Nanodrop ND-1000 spectrophotometer, using 2 µL samples of purified protein solution. The concentration was calculated from the absorption using molar extinction coefficient values of purified enzyme (Table 8.4) Enzyme coefficient values were calculated from protein sequences using the ProtParam tool on the ExPASy Proteomics Server(*195*).

Enzyme	Extinction coefficient
*Nme*NANAS wild-type*	22265
*Nme*G272Term	9315
*Cje*NANAS wild-type	26150
*Cje*LegS wild-type	102510
*Cje*PseS wild-type	23045
*Cte*SAS wild-type	44265
*Ssu*SAS wild-type	21890
Nme/*Hsa* chimera	16890

Table 8.4 Extinction coefficients of enzymes calculated using ProtParam (*195*). * The *Nme*NANAS wild-type extinction coefficient was used to determine the concentrations of *Nme*NANAS single and double point mutation variants as there was negligible difference in the calculated extinction coefficient values from ProtParam.

8.5 General characterisation procedures

8.5.1 Standard kinetic assay

The standard assay for SAS activity monitored the consumption of PEP by loss of absorbance at 232 nm, based on the method of Schoner and Herrmann (*196*). All kinetic measurements were made using a Varian Cary 100 UV-visible spectrophotometer, stoppered 1 cm pathlength quartz cuvettes and a total assay volume of 1 mL. Standard assays were carried out at 25 °C and all assay solutions were equilibrated to temperature and baseline absorbance before initialisation. The enzymatic reaction was initiated by the addition of 5 µL of purified enzyme at a stock concentration of 2-3 mg/mL. Standard reaction mixtures contained 50 mM BTP pH 7.5, a variable concentration of PEP, ManNAc, 1 mM $MnCl_2$ and enzyme.

Assay buffers were treated with Chelex® 100 resin (Bio-Rad) prior to use to remove metal ions, and Chelex-treated buffer solution was used to create concentrated stock solutions of PEP and ManNAc. Metal-ion solutions were created using either Chelex-treated buffer or Milli-Q water.

Initial rates of reaction were measured as a least-squares fit of the initial rate data using Cary WinUV Kinetics Application (version 3.00, Varian). A unit of enzyme activity was defined as the loss of 1 µmol of PEP min^{-1} at 25 °C and could be calculated from the measured loss of absorbance (at 232 nm) using Beer's Law ($\varepsilon = 2.8\times10^3$ M^{-1} cm^{-1} at 25 °C). The values of the kinetic parameters were determined by fitting data to the Michaelis-Menten equation $v = \frac{V_{max}[S]}{K_M+[S]}$ using the software GraFit 5(*197*) (Erithacus Software Limited). The concentration of one substrate was kept constant, while varying that of the other and vice versa. To determine apparent K_m (PEP), ManNAc concentration was held at 30 mM, and the concentration of PEP varied. To determine apparent K_m (ManNAc), PEP was kept constant at 1 mM, and ManNAc concentration was varied.

8.5.2 Mass spectrometry

The masses of purified proteins were measured by Bruker maXis 3G UHF-TOF tandem mass spectrometer. Proteins were either buffer exchanged into 5 mM ammonium bicarbonate, or were diluted from stored concentrations to 1 mg mL^{-1} with water.

8.5.3 Differential scanning fluorimetry

Melt temperatures of the proteins were determined both with and without the presence of different additives by differential scanning fluorimetry (DSF), using an iCycler iQ5 Multicolour Real-Time PCR Detection System (Bio-Rad). The method used was based on that of Ericsson *et al*.(*198*) Triplicate protein samples were added with mixing to buffer (potentially containing additives) and SYPRO orange dye (Invitrogen) in a 96-well microplate (*vide infra*), which was then sealed. The melt proceeded in 0.2 °C increments from 20 - 95 °C, with a 20 s dwell time after each temperature rise. Measurements of the fluorescence were made at the end of each dwell time.

For each sample (three replicates and blank) 100 µL of buffer (50 mM BTP pH 7.5) including any additive(s) was prepared, to which 5 µL of 250 x SYPRO orange dye was added (condition solution). In the blank well for the sample, 21 µL of the condition solution was added to 4 µL of water. Then, 16 µL of protein (at 1 mg mL^{-1}) was added to the remaining condition solution before 25 µL was dispensed into three replicate wells. Additives used in experiments were as follows: Metal (1 mM), PEP (1 mM) and ManNAc/rManNAc/ManNAc analogue (10 mM with metal, 1 mM with metal and PEP).

For analysis an Excel spreadsheet with custom VBA-scripted macro was created by Dr Timothy Allison to toggle and dynamically display melt and derivative curves for each sample. The melt temperatures were calculated as the temperature of maximum inflection of the melting curve after subtracting the reading of a blank well containing buffer and dye but lacking protein.

8.5.4 Circular dichroism

Circular dichroism (CD) experiments were performed using a JASCO J-815 Spectropolarimeter. All CD spectra were recorded from 260 nm to 195 nm using a 0.5 nm data pitch, 1 s response, 1 nm bandwidth at 25 °C. A 3 mL protein solution at a concentration of 0.03 mg mL^{-1} in a 3 mL quartz cuvette with 1 cm path length was used for each measurement. The buffer for all experiments was 20 mM sodium phosphate pH 8.0. Blank spectra were recorded of buffer alone for each experiment before addition of protein. Data were smoothed using the Savitzky-Golay algorithm of the Jasco Spectra Manager™ (version 1.5) with a convolution width of 15.

8.5.5 Isothermal calorimetry

The affinity of substrate and metal ions for enzyme was assessed by isothermal calorimetry (ITC) using a VP-ITC unit operating at 298 K (MicroCal; GE Healthcare). Prior to use, all solutions were filtered and degassed in a vacuum, and protein concentration was measured by UV absorption ($\varepsilon_{Nme\text{NANAS}}$ = 22,270 M^{-1} cm^{-1}). The binding buffer was composed of 30 mM triethanolamine (pH 7.5). To determine $K_{d(Mn)}$, 5 mM $MnCl_2$ was titrated into 70 µM enzyme. To determine $K_{d\ (PEP)}$, 50 mM PEP was titrated into 40 µM enzyme containing a background concentration of 1 mM $MnCl_2$. To determine $K_{d\ (rManNAc)}$, 10 mM rManNAc was titrated into 100 - 200 µM enzyme containing a background concentration of 1 mM $MnCl_2$ and 1 mM PEP. Titrations were performed with fifty-five injections of ligand; one 2 µL injection followed by fifty-four 5 µL injections. Heats of dilution experiments were measured independently and subtracted from the integrated data before curve-fitting in Origin 7.0 with the standard one-site model supplied by MicroCal.

8.5.6 Synthesis of rManNAc

Reduced ManNAc (rManNAc) was synthesised by Ryu Toyama using the method outlined in Gunawan et al (*112*). 0.21 g of ManNAc (1 mM) was dissolved in 10 mL of water and 0.08 g of sodium borohydride (2 mM) added to the solution. The mixture was subsequently stirred at room temperature for 2 hours. The resulting solution was neutralised using Dowex® 50WX8 hydrogen form, 100-200 mesh (Sigma-Aldrich). The solution was then filtered and lyophilised, producing a white solid. The product was subsequently refluxed in 50 mL of methanol for 30 min and concentrated under reduced pressure. The resultant oily film was re-dissolved in water and lyophilised to produce a white solid (0.15 – 0.16 g). Purity of rManNAc was determined through ^1H NMR to be above 99 %.

8.5.7 Analytical gel filtration chromatography

The multimeric states of enzyme were determined by gel filtration chromatography using a Superdex 200 10/300 GL column (GE Healthcare). Samples (500 µL, 1, 2 and 4 mg/mL) and standards (Sigma) were run in 50 mM bis-tris propane (pH 7.5) with 150 mM NaCl.

8.5.8 Analytical ultracentrifugation

Sedimentation velocity experiments were performed in a Beckman Coulter model XL-I analytical ultracentrifuge equipped with UV-visible scanning optics. Reference (400 µl; 30 mM triethanolamine, pH 7.5) and sample (380 µl) solutions were loaded into 12-mm double-sector cells with quartz windows, and the cells were then mounted in an An-50 Ti eight-hole rotor. Proteins were centrifuged at various rotor speeds at 20 °C, and radial absorbance data were collected at appropriate wavelengths in continuous mode every 8 min without averaging. Data were fitted to a continuous size-distribution (c_s) model using the program SEDFIT (*150, 151*). The partial specific volume ($\bar{v}$) of the proteins, buffer density (0.9988 g·ml^{-1}), and buffer viscosity (1.005 cp) were computed using the program SEDNTERP (*199*). Weight average molecular masses were calculated from weight average sedimentation coefficients and $f/f0$ values resulting from the fits using SEDFIT (*150, 151*).

8.5.9 Small angle X-ray scattering data collection

Measurements were performed at the Australian Synchrotron SAXS/WAXS beamline (where WAXS is wide angle X-ray scattering) equipped with a Pilatus detector (1 M, 170 X 170 mm, effective pixel size, 172 X 172 µm). The wavelength of the X-rays was 1.0332 Å. The sample detector distance was 1600 mm, which provided an s range of 0.0126 - 0.400Å^{-1} (where s is the magnitude of the scattering vector, which is related to the scattering angle (2Θ) and the wavelength (λ) as follows: $s = (4\pi/\lambda)\sin\Theta$). Scattering data was collected from enzymes at approximately 10 mg ml^{-1} following elution from a size-exclusion chromatography column (Superdex 200 5/150), pre-equilibrated with 30 mM Triethanolamine buffer at pH 7.5. As aforementioned *Cte*NANAS was unstable in triethanolamine and was thus eluted with 20 mM BTP and 150 mM NaCl at pH 7.5. Liganded SAXS samples contained PEP (1 mM) and Mn^{2+} (1 mM) in buffer. Two dimensional intensity plots from the peak of the size-exclusion chromatography run were radially averaged, normalized to sample transmission and background-subtracted.

8.5.10 Small angle X-ray scattering data analysis

Scattered intensity (*I*) was plotted *versus s.* Extrapolation of enzyme *I*(*s*) profiles to zero angle (*I*(0)) indicated a molecular mass consistent with the dimer molecular weights for each protein. All samples were devoid of an increase in intensity at low *s* (indicative of aggregation). Guinier plots were linear for $s \cdot Rg < 1.3$. The data sets for structural analyses were recorded with approximately 450 data points over the range $0.0135 \leq s \leq 0.4$ Å^{-1}. Indirect Fourier transform was performed using GNOM (*200*) to yield the function *P*(*r*), which gives both the relative probabilities of distances between scattering centers and the maximum dimension of the scattering particle D_{max}. Theoretical scattering curves were generated from atomic coordinates and compared with experimental scattering curves using CRYSOL (*153*).

To assess and compare the quality of the fits in CRYSOL, the χ_ν^2 values for the fits were compared as outlined in Mills *et al. (201)*.

8.5.11 Crystallisation

Crystals of variant *Nme*NANAS and *Ssu*SAS were grown by hanging-drop vapour diffusion in 24-well VDX plates (Hampton Research). Specific crystallisation protocols for *Nme*R314K and *Nme*R314A are outlined in Section 8.6.2 and for *Ssu*SAS in Section 8.9.2. Unless otherwise stated, the crystallisation trays were left at 20 °C until immediately before data collection, with crystals being transferred briefly into a cryoprotectant composed of 20% (v/v) ethylene glycol in the respective reservoir solution. Crystals typically began to form overnight for both *Nme*NANAS and *Ssu*SAS.

Crystal screens were either completed by hand or using a Mosquito Crystal® (TTP Labtech) crystallisation robot on a 96-well sitting drop iQ plate (TTP Labtech). Proteins were screened at varied concentrations between 5 and 15 mg/mL using commercially available crystallisation screens. Protein crystallisation conditions were screened using Molecular Dimensions Morpheus, Midas, JCSG-plus, PACT premier, and ProPlex HT-96 screens.

8.5.12 X-ray data collection

X-Ray diffraction data were collected at the Australian Synchrotron. Protein was transported frozen to Melbourne the day before collection. Crystals were mounted on loops, flash-frozen in liquid nitrogen and placed in a sample storage cassette (Crystal Positioning Systems), which was stored in a dry shipper (Taylor-Wharton CX100) and transported as checked-in luggage. Data collection at the Australian Synchrotron used either the Macromolecular Crystallography (MX1) or Micro Crystallography (MX2) beamlines(*202*) and were processed using iMosflm and Scala (CCP4 suite(*203*)).

8.5.13 Structure determination and refinement

Structures of *Nme*NANAS variant proteins and *Ssu*SAS were solved by molecular replacement, using the structure of wild-type *Nme*NANAS (PDB ID: 1XUZ) as a starting model (with all non-protein molecules removed) and performing ten rounds of rigid-body refinement using Refmac5.(*204*) The same set of reflections for calculation of R_{free} was carried through. Further refinements were conducted with Refmac5,(*204*) and electron-density maps were analysed with Coot.(*205*) The validation tools of Coot and MolProbity(*206*) were used to check for, and correct, conformation infelicities.

8.6 Methods in Chapter 2

8.6.1 Cloning

The *neuB* gene from *N. meningitidis* serotype B strain MC58 was amplified from genomic DNA using *Pfu*Ultra DNA polymerase (Agilent) as described in Section 8.3.2. Gene specific nested primers (Table 8.5) were designed incorporating generic extensions (underlined) for further rounds of nested PCR. The product from first round PCR was extended using generic gateway primers (Table 8.5) which incorporated the attb sites (italicised) The linear product from second round nested-PCR was cloned into the donor vector pDONR-221 using BP clonase enzyme mix as outlined in Section 8.3.8.

Oligonucleotide	Sequence (5' - 3')
*Nme*NANAS WT Fwd	GGCAGCGGCGGC**ATG**CAAAACAACAACGA
*Nme*NANAS WT Rev	GAAAGCTGGGTG**TTA**TTCAATATCAGT
Generic Nested Fwd	GGGG*ACAAGTTTGTACAAAAAAGCAGGCTT*C**GAAAACCTGTATTTTCAGGGC**AGCGGCGCG
Generic Nested Rev	GGGGA*CCACTTTGTACAAGAAAGCTGGGT*

Table 8.5 First and second round nested primers used for amplification of *N. meningitidis* NANAS from genomic DNA. Start codons are in blue and stop codons are in red. Overlapping sequences for generic primer extension are underlined. TEV protease cleavage site is in bold and the attB1 (Fwd) and attb2 (Rev) sites are italicised.

Variants of *Nme*NANAS were generated by site-directed mutagenesis. The primers displayed in Table 8.6 were used to generate *Nme*R314K and *Nme*R314A, using the aforementioned pDEST-17-*Nme*NANAS plasmid as a template.

Oligonucleotide	Sequence (5' - 3')
*Nme*R314K Fwd	GAGATAACCTATGGGTTAAAAAACCAGGCAATGGAGACTTCAG
*Nme*R314K Rev	CTGAAGTCTCCATTGCCTGGTTTTTAACCCATAGGTTATCTC
*Nme*R314A Fwd	GATAACCTATGGGTTAAAGCGCCAGGCAATGGAGACTTC
*Nme*R314A Rev	GAAGTCTCCATTGCCTGGCGCTTTAACCCATAGGTTATC

Table 8.6 Mutagenesis primers used for amplification of *Nme*R314K and *Nme*R314A from pDEST17-*Nme*NANAS. Mutation sites are underlined.

8.6.2 Crystallisation of *Nme*R314K and *Nme*R314A

Purified R314K at a concentration of 12 mg/mL was crystallised in 1.9 M malate and 10 mM $MnCl_2$ at pH 6.2, and R314A at a concentration of 11.5 mg/mL was crystallised in 1.7 M malate and 10 mM $MnCl_2$ at pH 6.2 using hanging drop vapor diffusion. Hanging drops were 2 µL in size combining 1 µL of protein with 1 µL of reservoir solution and the total volume of the reservoir solution was 500 µL. Data sets were collected at the Australian Synchrotron using the MX1 and MX2 beamlines. Orthorhombic crystals with a space group of $P2_12_12$ were obtained of both mutants. The structures for both R314K and R314A were solved by molecular replacement using the structure of the wild-type protein (1XUZ) (*112*). Measured intensities were converted into amplitudes by using TRUNCATE as part of Scalepack2mtz or dtrek2mtz (CCP4) (*203*). The structure was then solved by performing two rounds of rigid-body refinement (CCP4). Intensity-based refinement was carried out with REFMAC5 (CCP4) following each cycle of structure development from Coot. Water molecules were added automatically following the criterion that they had at least one hydrogen bonding partner to a protein atom and/or water molecule in the range 2.4–3.2 Å. Atomic coordinates and structure amplitudes have been deposited with the PDB under accession codes 4IPJ for R314K and 4IPI for R314A.

8.6.3 Modelling of substrates in *Nme*R314K and *Nme*R314A

Modelling studies were conducted by Dr Wanting Jiao, using software packages from Schrodinger Suite 2011. The structures of the substrate ManNAc were built in Maestro (*19*) and then prepared using LigPrep (*20*). Crystal structures of the wild-type (PDB code 1XUZ), were used as the wild-type enzyme for docking. Although the electron density map for the bound rManNAc in this crystal structure is not great, rManNAc would be removed prior to docking calculation, so that it does not affect the result of the modelling study. The wild-type crystal structure was then prepared using Protein Preparation Wizard in Schrodinger Suite 2011 (*21*). Due to the malate molecules bound in the active site of R314A and R314K crystal structures, several active site residues adopted slightly different conformations to those observed in the wild-type crystal structure, e.g. the active site metal ions of R314A and R314K were shifted in position due to malate molecules coordinating to the metal ions; the residue Tyr186 also showed alternative conformations due to the flexibility of the active site residues in the absence of substrates PEP and ManNAc.

Since these residues all have interactions to the substrate bound in the active site, small changes in their positions may affect the outcome of a modelling study, and the differences seen in the modelling results for wild-type and the mutant systems may not reflect only the effect of mutation. Therefore, the enzyme structures of R314K and R314A mutants were made by *in silico* mutation of Arg314 in the prepared wild-type crystal structure (PDB code 1XUZ) to the corresponding Lys and Ala. All other residues in the wild-type structure including the substrate PEP in the active site were retained in the *in-silico*-generated mutant enzymes.

The mutant enzymes from *in silico* mutagenesis showed little difference in side-chain and backbone positions at residue 314 compared to those in the corresponding crystal structures. For all three enzymes, the center of the receptor grid was defined as the centroid of the rManNAc molecule bound in the active site of Chain A; The docking of flexible ManNAc molecule into rigid enzyme receptors of wild-type, R314A and R314K enzymes was conducted in Glide (*22, 207-209*) with OPLS2005 force field and extra precision (XP) mode. The van der Waals radii of ligand atoms were not scaled. 90000 poses per ligand were kept for initial docking, and scoring window for keeping initial poses was 5000. The best 1000 poses per ligand were kept for energy minimization with a distance dependent dielectric constant of 2 and a maximum of 5000 conjugate gradient steps.

Before any output poses were written out, the poses obtained in the docking calculation were clustered and duplicate poses were discarded if the RMS deviation is less than 0.5 Å and maximum atomic displacement is less than 1.3 Å. This clustering procedure ensures that the output poses were conformationally distinct, as a result different number of output poses was usually written out for different sets of calculations.

In order to assess how closely the output poses from the modelling study resembles the crystallographically observed binding pose of the substrate analogue rManNAc, RMS deviations between poses from modelling study and the crystal structure were measured using the measurement tool in Maestro (*19*). The wild-type and mutant receptors used for the modelling study were first superimposed with the wild-type crystal structure, and then the RMS deviations between the modelled poses and the crystal structure of rManNAc were measured in place. Since there are no hydrogen atoms in the original crystal structure, hydrogen atoms in the modelled poses were not included for the calculation of RMS deviations.

8.7 Methods in Chapter 3

8.7.1 Cloning

Variants of *Nme*NANAS were generated by site-directed mutagenesis using *Pfu*Ultra DNA polymerase (Agilent) as described in Section 8.3.2. The truncation G272Term was made by inserting a stop codon. The following primers (Table 8.7) were used to generate mutants using the aforementioned pDEST17-*Nme*NANAS plasmid as a template. The double mutation E282A/T285A used pDEST17-*Nme*NANAS E282A as a template.

Oligonucleotide	Sequence (5' - 3')
*Nme*G272Term Fwd	CTCATGCTTTAAAATTGGCACGCTAGGGCAAAAAAGACACGATTATCGC
*Nme*G272Term Rev	GCGATAATCGTGTCTTTTTTGCCCTAGCGTGCCAATTTTAAAGCATGAG
*Nme*T285A Fwd	TCGCGGGAGAAAAGCCAGCTAAAGATTTCGCCTTT
*Nme*T285A Rev	AAAGGCGAAATCTTTAGCTGGCTTTTCTCCCGCGA
Nme T285F Fwd	ATTATCGCGGGAGAAAAGCCATTTAAAGATTTCGCCTTTGCATC
Nme T285F Rev	GATGCAAAGGCGAAATCTTTAAATGGCTTTTCTCCCGCGATAAT
*Nme*E282A Fwd	GACACGATTATCGCGGGAGCAAAGCCAACTAAAGATTTC
*Nme*E282A Rev	GAAATCTTTAGTTGGCTTTGCTCCCGCGATAATCGTGTC
*Nme*E282A/T285A Fwd	CACGATTATCGCGGGAGCAAAGCCAGCTAAAGATTTCGCCTTT
*Nme*E282A/T285A Rev	AAAGGCGAAATCTTTAGCTGGCTTTGCTCCCGCGATAATCGTG
*Nme*E134A Fwd	GCATATAAAATCGGCTCTGGCGCATGTAATAACTACCCATTAATT
*Nme*E134A Rev	AATTAATGGGTAGTTATTACATGCGCCAGAGCCGATTTTATATGC

Table 8.7 Mutagenesis primers used for amplification of *Nme*G272Term, *Nme*T285A, *Nme*T285F, *Nme*E282A, *Nme*E282A/T285A and *Nme*E134A. Mutation sites are underlined.

8.7.2 Molecular dynamics simulations

Molecular dynamics simulations were conducted with (*210*) running on the BlueFern supercomputer at the University of Canterbury, New Zealand. Prior to molecular dynamics simulation, crystallographic water and ligands were deleted from the *Nme*NANAS crystal structure (PDB IXUZ). Each starting structure was solvated with explicit TIP3 water molecules in a box in (*211*), and ionized by adding Na+ and Cl- ions to balance the net charge of the water box. Ions were added with a minimum distance of 5 Å from the molecule and from each other. Topology and parameter files for PEP molecule was generated with SwissParam online server (http://swissparam.ch/) based on mol2 file obtained from ZINC database. Molecular dynamics simulations were conducted with the (*212*) all-hydrogen parameter file for proteins at constant temperature (322.15 K) and pressure (1 atm). The cut-off distance for van der Waals interactions was set to 12 Å. In each simulation, the system was first minimized for 5000 steps followed by 270 ns of dynamics simulations conducted with 2 fs time steps. The trajectory was written out at 100 ps time intervals, and a total of 2700 frames were obtained from each simulation.

8.7.3 PISA

Macromolecular interfaces were assessed using the online PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) (*152*). Differences in surface area contacts for wild-type *Nme*NANAS and *Nme*G272Term were determined using the PDB coordinates *Nme*NANAS 1XUZ (*112*). PDB coordinates 1XUZ were modified to generate a model of *Nme*G272Term which lacked the AFPL and linker domains. PISA was used to calculate the interdomain surface area, number of interfacing atoms ($^{i}N_{at}$) as well as the solvation free energy gain yielded upon interface formation ($\Delta^{i}G$).

8.8 Methods in Chapter 4

8.8.1 Cloning

The genes encoding NANAS, LegS and PseS from *C. jejuni* NCTC 11168 were amplified from genomic DNA using *Pfu*Ultra DNA polymerase (Agilent) as described in Section 8.3.2. Gene specific nested primers (Table 8.8) were designed incorporating generic extensions (underlined) for further rounds of nested PCR. The product from first round PCR was extended using generic gateway primers (Table 8.8) which incorporated the attb sites (italicised) The linear product from second round nested-PCR was cloned into the donor vector pDONR-221 using BP clonase enzyme mix as outlined in Section 8.3.8.

Oligonucleotide	Sequence (5' - 3')
*Cje*NANAS (Cj1141) Fwd	GGCAGCGGCGCG**ATG**CAAATAAAAATAGATA
*Cje*NANAS (Cj1141) Rev	GAAAGCTGGGTG**TCA**TTCAAAATCATCCCATG
*Cje*LegS (Cj1327) Fwd	GGCAGCGGCGCG**ATG**AAAAAAACTTTAATCAT
*Cje*LegS (Cj1327) Rev	GAAAGCTGGGTG**TTA**CTCACGGATAAGCTCATCTTCT
*Cje*PseS (Cj1317) Fwd	GGCAGCGGCGCG**ATG**CAAATAGGAAATTTTAACACCG
*Cje*PseS (Cj1317) Rev	GAAAGCTGGGTG**TCA**TTGGAAATCTCCTTGTTTTAAA
Generic Nested Fwd	GGGG*ACAAGTTTGTACAAAAAAGCAGGCT*TC**GAAAACCTGTATTTTCAGGGC**AGCGGCGCG
Generic Nested Rev	GGGGA*CCACTTTGTACAAGAAAGCTGGGT*

Table 8.8 First and second round nested primers used for amplification of *C. jejuni* SASs from genomic DNA. Start codons are in blue and stop codons are in red. Overlapping sequences for generic primer extension are underlined. TEV protease cleavage site is in bold and the attB1 (Fwd) and attb2 (Rev) sites are italicised.

8.8.2 Phylogram sequence selection

The sequences of NANAS, LegS, PseS and putative SASs were retrieved from KEGG (*213*). In this database, several of the genes this thesis considers to be LegSs, PseSs or Type IV SASs were annotated as NANASs or putative NANASs.

Very few sequences in KEGG were annotated as LegS or PseS and none were annotated as Type IV SASs as this notation is an invention of this thesis. In order to attain more LegS, PseS and Type IV sequences a protein BLAST (*164*) search was performed using *Cje*LegS, *Cje*PseS and *Cte*SAS as queries. From this BLAST search other putative LegS, PseS and Type IV SASs were selected for sequence alignment. Each of these sequences had at least 30 % sequence identity to the queried sequence.

8.8.3 Modelling of *Cje*NANAS, *Cje*LegS and *Cje*PseS

Since crystal structures of *Cje*NANAS, *Cje*LegS and *Cje*PseS could not be obtained, models were generated from the SWISS-MODEL server (http://swissmodel.expasy.org/) (*162*) using wild-type *Nme*NANAS PDB coordinates 1XUZ (*112*) as a template. The sequences of *Cje*NANAS, *Cje*LegS and *Cje*PseS were retrieved from the KEGG (*213*) database and aligned with the sequence of *Nme*NANAS using ClustalW (*155*). The sequence identity of *Cje*NANAS, *Cje*PseS and *Cje*LegS with *Nme*NANAS was determined to be 58.6 %, 32.4 % and 34.1 % respectively.

Model quality was verified using online SWISS-MODEL assessment software. QMEAN composite scoring function was used to determine whether the models were correctly folded. QMEAN considers torsion angles, interaction potentials, solvation potential and the agreement of predicted secondary structure (from sequence) with calculated secondary structure. Models with QMEAN scores closest to 0 are deemed more reliable. The QMEAN scores determined for *Cje*NANAS, *Cje*PseS and *Cje*LegS were -0.23, -0.90 and -0.66 respectively.

8.8.4 Substrate docking of *Cje*NANAS, *Cje*LegS and *Cje*PseS

Induced fit docking of Mn^{2+}, PEP and 2, 4-diacetamido-2, 4, 6-trideoxyaltrose into the model of *Cje*PseS and induced fit docking of Mn^{2+}, PEP and 2, 4-diacetamido-2, 4, 6-trideoxymannose into the model of *Cje*LegS were carried out by Dr Wanting Jiao. Induced fit docking was completed using Schrödinger Suite Software (Maestro, LigPrep, Epik, Glide) (*207-209*).

8.9 Methods in Chapter 5

8.9.1 Crystallisation of *Ssu*SAS

Purified *Ssu*SAS at a concentration of 6 mg/mL was crystallised using a condition obtained from a Morpheus™ crystallisation screen (Molecular Dimensions). Apoenzyme *Ssu*SAS crystals were formed using the hanging drop crystallisation technique in the presence of 0.09 M $NaNO_3$/Na_2HPO_4/$NH_4)_2SO_4$, 0.1 M Imidazole/MES and 37.5% racemic MPD/PEG 1000/PEG 3350 at pH 6.5. Hanging drops were 2 µL in size combining 1 µL of protein with 1 µL of reservoir solution and the total volume of the reservoir solution was 500 µL. Data sets were collected at the Australian Synchrotron using the MX1 and MX2 beamlines.

The structure of *Ssu*SAS was solved by molecular replacement using the structure of the wild-type protein (1XUZ) (*112*). Measured intensities were converted into amplitudes by using TRUNCATE as part of Scalepack2mtz or dtrek2mtz (CCP4) (*203*). The structure was then solved by performing ten rounds of rigid-body refinement (CCP4). Intensity-based refinement was carried out with REFMAC5 (CCP4) following each cycle of structure development from Coot. Water molecules were added automatically following the criterion that they had at least one hydrogen bonding partner to a protein atom and/or water molecule in the range 2.4–3.2 Å.

8.9.2 Substrate docking in *Ssu*SAS

Induced fit docking analyses of Mn^{2+}, PEP and ManNAc/2, 4-diacetamido-2, 4, 6-trideoxymannose into the active site of *Ssu*SAS were carried out by Dr Wanting Jiao. Induced fit docking was completed using Schrödinger Suite Software (Maestro, LigPrep, Epik, Glide) (*207-209*).

8.10 Methods in Chapter 6

8.10.1 Cloning

The constructs for the *Hsa*/*Nme* chimera and *Nme*/*Has* chimera were generated from the genes encoding *Nme*NANAS and *Hsa*NANA-9-PS using a multi-step PCR protocol that is outlined in Section 6.3. The chimeric constructs were cloned using *Pfu*Ultra DNA polymerase (Agilent) as described in Section 8.3.2. The following primers (Table 8.9) were used to generate the *Hsa*/*Nme* chimera and *Nme*/*Hsa* chimera using pDEST17-*Nme*NANAS as well as a pDEST17-*Hsa*NANA-9-PS as templates.

Oligonucleotide	Sequence (5' - 3')
*Nme*NANAS catalytic domain Fwd	GGCAGCGGCGCGATGCAAAACAACAACGAATTT
*Hsa*NANA-9-PS AFPL domain Rev	GAAAGCTGGGTGTTAGCTTTTGATTTTTTTGCC
Nme/*Hsa* Chimera bridge Fwd	GCTTTAAAATTGGCACGCGGCAGCCCGACCAAACAGCTGCTG
Nme/*Hsa* Chimera bridge Rev	CAGCAGCTGTTTGGTCGGGCTGCCGCGTGCCAATTTTAAAGC
*Hsa*NANA-9-PS catalytic domain Fwd	GGCAGCGGCGCGATGCCGCTGGAACTGGAACTG
*Nme*NANAS AFPL domain Rev	GAAAGCTGGGTGTTATTCAATATCAGTTTTTTTG
Hsa/*Nme* Chimera bridge Fwd	CTGGTTGAACGTGCCCTGGGTGGCAAAAAAGACACGATTATC
Hsa/*Nme* Chimera bridge Rev	GATAATCGTGTCTTTTTTGCCACCCAGGGCACGTTCAACCAG

Table 8.9 Primers used for generation of *Nme*/*Hsa and Hsa*/*Nme* chimeras. Start codons are in blue and stop codons are in red. Overlapping sequences for generic primer extension are underlined. Sequence corresponding to the *Nme*NANAS catalytic domain is highlighted in pink, *Nme*NANAS AFPL domain in green, *Hsa*NANA-9-PS catalytic domain in blue and *Hsa*NANA-9-PS AFPL domain in green.

8.10.2 *Hsa*NANA-9-PS cloning and purification

The gene encoding *Hsa*NANA-9-PS was codon optimised for expression in an *E.coli* cell line, synthesised and cloned into the vector pDONR 221 by GeneArt®. DNA sequence encoding an N-terminal TEV protease cleavage site was incorporated during the construction of the *Hsa*NANA-9-PS gene. The *Hsa*NANA-9-PS gene was sub-cloned into the expression vector pDEST-17 which encodes an N-terminal His-tag by Thomas Cotton. Expression trials and purification of *Hsa*NANA-9-PS were carried out Thomas cotton using the methods outlined in Section 8.4.

8.10.3 *Hsa*NANA-9-PS modelling

Homology modelling of *Hsa*NANA-9-PS was completed by Dr Wanting Jiao. Since no full-length crystal structure of *Hsa*NANA-9-PS is currently available, a homology model was generated based on the *Hsa*NANA-9-PS amino acid sequence, using the crystal structure of *Nme*NANAS PDB code 1XUZ (*112*) as a template. Sequence alignment of *Hsa*NANA-9-PS and *Nme*NANAS was generated using the ClustalW online server (*155*). The resulting alignment was used to generate the homology model, using Maestro 9.3.5 (Schrödinger Software). The model output was refined following alignment with the solved solution structure of *Hsa*NANA-9-PS AFP domain (PDB code 1WVO).

Induced fit docking of ManNAc-6-P into active site of the homology model of *Hsa*NANA-9-PS was conducted for both monomers present in the dimer of the homology model, due to the difference in conformation of the predicted ManNAc binding loops in both monomers. Induced fit docking was completed using Schrödinger Suite Software (Maestro, LigPrep, Epik, Glide) (*207-209*).

References

1. Schauer, R. (2004) Sialic acids: fascinating sugars in higher animals and man, *Zoology 107*, 49-64.
2. Traving, C., and Schauer, R. (1998) Structure, function and metabolism of sialic acids, *Cell. Mol. Life Sci. 54*, 1330-1349.
3. Schauer, R. (1985) Sialic acids and their role as biological masks, *Trends Biochem.Sci. 10*, 357-360.
4. Angata, T., and Varki, A. (2002) Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective, *Chem. Rev. 102*, 439-469.
5. Tanner, M. E. (2005) The enzymes of sialic acid biosynthesis, *Bioorganic Chem. 33*, 216-228.
6. Warren, L. (1959) Thiobarbituric acid assay of sialic acids, *J. Biol. Chem. 234*, 1971-1975.
7. Zhang, Y. N., Tao, F., Du, M. F., Ma, C. Q., Qiu, J. H., Gu, L. C., He, X. F., and Xu, P. (2010) An efficient method for *N*-acetyl-D-neuraminic acid production using coupled bacterial cells with a safe temperature-induced system, *Appl. Microbiol. Biotechnol. 86*, 481-489.
8. Brinkman-Van der Linden, E. C. M., Sjoberg, E. R., Juneja, L. R., Crocker, P. R., Varki, N., and Varki, A. (2000) Loss of *N*-glycolylneuraminic acid in human evolution - Implications for sialic acid recognition by siglecs, *J. Biol. Chem. 275*, 8633-8640.
9. Kawano, T., Kozutsumi, Y., Takematsu, H., Kawasaki, T., and Suzuki, A. (1993) Regulation of biosynthesis of *N*-glycolylneuraminic acid containing glycoconjugates - Characterisation of factors required for NADH-dependent cytidine 5'-monophosphate-*N*-acetylneuraminic acid hydroxylation, *Glycoconjugate J. 10*, 109-115.
10. Malykh, Y. N., Schauer, R., and Shaw, L. (2001) *N*-glycolylneuraminic acid in human tumours, *Biochimie 83*, 623-634.
11. Angata, T., Nakata, D., Matsuda, T., Kitajima, K., and Troy, F. A. (1999) Biosynthesis of KDN (2-keto-3-deoxy-D-glycero-D-galacto-nononic acid) - Identification and characterization of a KDN-9-phosphate synthetase activity from trout testis, *J. Biol. Chem. 274*, 22949-22956.
12. Inoue, S., and Kitajima, K. (2006) KDN (deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family, *Glycoconj. J. 23*, 277-290.
13. Chen, X., and Varki, A. (2010) Advances in the biology and chemistry of sialic acids, *Acs Chem. Bio. 5*, 163-176.
14. Alviano, C. S., Travassos, L. R., and Schauer, R. (1999) Sialic acids in fungi: A minireview, *Glycoconjugate Journal 16*, 545-554.

15. Bakker, H., Routier, F., Ashikov, A., Neumann, D., Bosch, D., and Gerardy-Schahn, R. (2008) A CMP-sialic acid transporter cloned from *Arabidopsis thaliana*, *Carbohydr. Res. 343*, 2148-2152.

16. Takashima, S., Seino, J., Nakano, T., Fujiyama, K., Tsujimoto, M., Ishida, N., and Hashimoto, Y. (2009) Analysis of CMP-sialic acid transporter-like proteins in plants, *Phytochemistry 70*, 1973-1981.

17. Zeleny, R., Kolarich, D., Strasser, R., and Altmann, F. (2006) Sialic acid concentrations in plants are in the range of inadvertent contamination, *Planta 224*, 222-227.

18. Onodera, K., Hirano, S., and Hayashi, H. (1966) Sialic acid and related substances 4. Sialic acid content of some biological materials, *Ag. Biol. Chem. 30*, 1170-&.

19. Harduin-Lepers, A., Mollicone, R., Delannoy, P., and Oriol, R. (2005) The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach, *Glycobiology 15*, 805-817.

20. Zhao, W., Chen, T. L. L., Vertel, B. M., and Colley, K. J. (2006) The CMP-sialic acid transporter is localized in the medial-trans Golgi and possesses two specific endoplasmic reticulum export motifs in its carboxyl-terminal cytoplasmic tail, *J. Biol. Chem. 281*, 31106-31118.

21. Alviano, C. S., Pereira, M. E. A., Souza, W., Oda, L. M., and Travassos, L. R. (1982) Sialic acids are surface components of *Sporothrix schenckii* yeast forms, *Fems Microbiol. Lett. 15*, 223-227.

22. Benhamou, N., and Ouellete, G. B. (1987) Ultrastructural characterisation of an extracellular fibrillar sheath on cells of *Ascocalyx abietina*, the scleroderris canker agent of conifers, *Can. J. Bot. Rev. 65*, 154-167.

23. Hamilton, A. J., Jeavons, L., Hobby, P., and Hay, R. J. (1992) A 34-kilodalton to 38-kilodalton *Cryptococcus neoformans* glycoprotein produced as an exoantigen bearing a glycosylated species specific epitope, *Infect. Immun. 60*, 143-149.

24. Souza, E. T., Silvafilho, F. C., Desouza, W., Alviano, C. S., Angluster, J., and Travassos, L. R. (1986) Identification of sialic acids on the cell-surface of hyphae and conidia of the human pathogen *Fonsecae pedrosoi*, *J. Med. Vet. Mycol. 24*, 145-154.

25. Soares, R. M. A., Silva, F. C., Rozental, S., Angluster, J., de Souza, W., Alviano, C. S., and Travassos, L. R. (1998) Anionogenic groups and surface sialoglycoconjugate structures of yeast forms of the human pathogen *Paracoccidioides brasiliensis*, *Microbiol. U. K. 144*, 309-314.

26. Alviano, C. S., Travassos, L. R., and Schauer, R. (1999) Sialic acids in fungi: A minireview, *Glycocon. J. 16*, 545-554.

27. Roth, J., Kempf, A., Reuter, G., Schauer, R., and Gehring, W. J. (1992) Occurence of sialic acids in *Drosophila melanogaster*, *Science 256*, 673-675.
28. Schauer, R. (2001) The occurrence and significance of sialic acids in insects, *Trends Glycosci. Glycotechnol. 13*, 507-517.
29. Viswanathan, K., Tomiya, N., Park, J., Singh, S., Lee, Y. C., Palter, K., and Betenbaugh, M. J. (2006) Expression of a functional *Drosophila melanogaster* CMP-sialic acid synthetase - Differential localization of the drosophila and human enzymes, *J. Biol. Chem. 281*, 15929-15940.
30. Kim, K., Lawrence, S. M., Park, J., Pitts, L., Vann, W. F., Betenbaugh, M. J., and Palter, K. B. (2002) Expression of a functional *Drosophila melanogaster N*-acetylneuraminic acid (Neu5Ac) phosphate synthase gene: evidence for endogenous sialic acid biosynthetic ability in insects, *Glycobiology 12*, 73-83.
31. Karacali, S., Kirmizigul, S., and Deveci, R. (1999) Sialic acids in developing testis of *Galleria mellonella* (*Lepidoptera*), *Invert. Rep. Dev. 35*, 225-229.
32. Kyriakides, T. R., McKillip, J. L., and Spence, K. D. (1995) Biochemical characterisation, developmental expression, and induction of the immune protein scolexin from *Manduca sexta*, *Arc. Ins. Biochem. Phys. 29*, 269-280.
33. Bidochka, M. J., StLeger, R. J., and Roberts, D. W. (1997) Induction of novel proteins in *Manduca sexta* and *Blaberus giganteus* as a response to fungal challenge, *J. Invertebr. Pathol. 70*, 184-189.
34. Palomares, L. A., Joosten, C. E., Hughes, P. R., Granados, R. R., and Shuler, M. L. (2003) Novel insect cell line capable of complex *N*-glycosylation and sialylation of recombinant proteins, *Biotech. Prog. 19*, 185-192.
35. Okazaki, T., Okudaira, N., Iwabuchi, K., Fugo, H., and Nagai, T. (2006) Apoptosis and adhesion of hemocytes during molting stage of silkworm, *Bombyx mori*, *Zool. Sci. 23*, 299-304.
36. Schauer, R. (2000) Achievements and challenges of sialic acid research, *Glyco. J. 17*, 485-499.
37. Varki, A. (1997) Sialic acids as ligands in recognition phenomena, *Faseb J. 11*, 248-255.
38. Baron, S. (1996) *Medical microbiology* Vol. 4th edition.
39. Ferrari, G., Batistatou, A., and Greene, L. A. (1993) Gangliosides rescue neuronal cells from death after trophic factor deprivation, *J. Neurosci. 13*, 1879-1887.
40. Wang, B., Miller, J. B., McNeil, Y., and McVeagh, P. (1998) Sialic acid concentration of brain gangliosides: Variation among eight mammalian species, *Comp. Biochem. Physiol. A-Mol. Integr. Physiol. 119*, 435-439.

41. Modi, P., Sadasivudu, B., Lakshminarayana, U., and Murthy, C. R. K. (1994) Functional relationship between ammonia and gangliosides in brain, *Neurochem Res 19*, 353-358.
42. Spichtig, V., Michaud, J., and Austin, S. (2010) Determination of sialic acids in milks and milk-based products, *Anal. Biochem. 405*, 28-40.
43. Baos, S. C., Phillips, D. B., Wildling, L., McMaster, T. J., and Berry, M. (2012) Distribution of sialic acids on mucins and gels: A defense mechanism, *Biophys. J. 102*, 176-184.
44. Yamaguchi, T., Uno, T., Uekusa, Y., Yagi-Utsumi, M., and Kato, K. (2013) Ganglioside-embedding small bicelles for probing membrane-landing processes of intrinsically disordered proteins, *Chem. Comm. 49*, 1235-1237.
45. Wang, B. (2009) Sialic acid is an essential nutrient for brain development and cognition, In *Annu. Rev. Nutr.*, pp 177-222.
46. Schwarzkopf, M., Knobeloch, K. P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R. (2002) Sialylation is essential for early development in mice, *Proc. Natl. Acad. Sci. U. S. A. 99*, 5267-5270.
47. Munday, J., Floyd, H., and Crocker, P. R. (1999) Sialic acid binding receptors (siglecs) expressed by macrophages, *J. Leuk. Biol. 66*, 705-711.
48. Crocker, P. R., Mucklow, S., Bouckson, V., McWilliam, A., Willis, A. C., Gordon, S., Milon, G., Kelm, S., and Bradfield, P. (1994) Sialoadhesin, a macrophage sialic acid binding receptor for hematopoietic cells with 17 immunoglobulin-like domains, *Embo J. 13*, 4490-4503.
49. Kelm, S., Pelz, A., Schauer, R., Filbin, M. T., Tang, S., Debellard, M. E., Schnaar, R. L., Mahoney, J. A., Hartnell, A., Bradfield, P., and Crocker, P. R. (1994) Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid dependent adhesion molecules of the immunoglobulin superfamily, *Curr. Biol. 4*, 965-972.
50. Varki, N. M., and Varki, A. (2007) Diversity in cell surface sialic acid presentations: implications for biology and disease, *Lab. Invest. 87*, 851-857.
51. Fuster, M. M., and Esko, J. D. (2005) The sweet and sour of cancer: Glycans as novel therapeutic targets, *Nature Reviews Cancer 5*, 526-542.
52. Seidenfaden, R., Krauter, A., Schertzinger, F., Gerardy-Schahn, R., and Hildebrandt, H. (2003) Polysialic acid directs tumor cell growth by controlling heterophilic neural cell adhesion molecule interactions, *Molecular and Cellular Biology 23*, 5908-5918.
53. Fleiszig, S. M. J., Zaidi, T. S., Ramphal, R., and Pier, G. B. (1994) Modulation of *Pseudomonas aeruginosa* adherence to the corneal surface by mucus, *Infect. Immun. 62*, 1799-1804.
54. Varghese, J. N. (1999) Development of neuraminidase inhibitors as anti-influenza virus drugs, *Drug Dev. Res. 46*, 176-196.

55. Moscona, A. (2005) Drug therapy - Neuraminidase inhibitors for influenza, *N. Engl. J. Med. 353*, 1363-1373.

56. Achyuthan, K. E., and Achyuthan, A. M. (2001) Comparative enzymology, biochemistry and pathophysiology of human exo-alpha-sialidases (neuraminidases), *Comp. Biochem. Physiol. B-Biochem. Mol. Biol. 129*, 29-64.

57. Shtyrya, Y. A., Mochalova, L. V., and Bovin, N. V. (2009) Influenza virus neuraminidase: Structure and function, *Acta Naturae 1*, 26-32.

58. Allen, S., Zaleski, A., Johnston, J. W., Gibson, B. W., and Apicella, M. A. (2005) Novel sialic acid transporter of *Haemophilus influenzae*, *Infect. Immun. 73*, 5291-5300.

59. Severi, E., Hood, D. W., and Thomas, G. H. (2007) Sialic acid utilization by bacterial pathogens, *Microbiology-Sgm 153*, 2817-2822.

60. Preston, A., Mandrell, R. E., Gibson, B. W., and Apicella, M. A. (1996) The lipooligosaccharides of pathogenic Gram-negative bacteria, *Crit. Rev. Microbiol. 22*, 139-180.

61. Knirel, Y. A., Kocharova, N. A., Shashkov, A. S., Kochetkov, N. K., Mamontova, V. A., and Soloveva, T. F. (1989) Structure of the capsular polysaccharide of *Klebsiella ozaenae* serotype K4 containing 3-deoxy-D-glycero-D-galacto-nonulosonic acid, *Carbohyd. Res. 188*, 145-155.

62. Josenhans, C., Labigne, A., and Suerbaum, S. (1995) Comparative ultrastructural and functional studies of the *Helicobacter pylori* and *Helicobacter mustelae* flagellin mutants - Both flagellin subunits, FLAA and FLAB, are necessary for full motility in *Helicobacter* species *J. Bact. 177*, 3010-3020.

63. Schoenhofen, I. C., McNally, D. J., Brisson, J. R., and Logan, S. M. (2006) Elucidation of the CMP-pseudaminic acid pathway in Helicobacter pylori: synthesis from UDP-*N*-acetylglucosamine by a single enzymatic reaction, *Glycobiology 16*, 8C-14C.

64. Linton, D., Karlyshev, A. V., Hitchen, P. G., Morris, H. R., Dell, A., Gregson, N. A., and Wren, B. W. (2000) Multiple N-acetyl neuraminic acid synthetase (neuB) genes in Campylobacter jejuni: identification and characterization of the gene involved in sialylation of lipo-oligosaccharide, *Molecular Microbiology 35*, 1120-1134.

65. Vale, R. D. (2003) The molecular motor toolbox for intracellular transport, *Cell 112*, 467-480.

66. Berg, H. C. (2003) The rotary motor of bacterial flagella, *Annu. Rev. Biochem. 72*, 19-54.

67. Thibault, P., Logan, S. M., Kelly, J. F., Brisson, J. R., Ewing, C. P., Trust, T. J., and Guerry, P. (2001) Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin, *J. Biol. Chem. 276*, 34862-34870.

68. Glaze, P. A., Watson, D. C., Young, N. M., and Tanner, M. E. (2008) Biosynthesis of CMP-*N*,*N*'-diacetyllegionaminic acid from UDP-*N*,*N*'-diacetylbacillosamine in *Legionella pneumophila*, *Biochemistry 47*, 3272-3282.

69. Fields, B. S., Benson, R. F., and Besser, R. E. (2002) Legionella and Legionnaires' disease: 25 years of investigation, *Clinic. Microbio. Rev. 15*, 506-526.

70. Albert-Weissenberger, C., Cazalet, C., and Buchrieser, C. (2007) *Legionella pneumophila* - a human pathogen that co-evolved with fresh water protozoa, *Cell. Mol. Life Sci. 64*, 432-448.

71. McNally, D. J., Aubry, A. J., Hui, J. P. M., Khieu, N. H., Whitfield, D., Ewing, C. P., Guerry, P., Brisson, J. R., Logan, S. M., and Soo, E. C. (2007) Targeted metabolomics analysis of Campylobacter coli VC167 reveals legionaminic acid derivatives as novel flagellar glycans, *J. Biol. Chem. 282*, 14463-14475.

72. Howard, S. L., Jagannathan, A., Soo, E. C., Hui, J. P. M., Aubry, A. J., Ahmed, I., Karlyshev, A., Kelly, J. F., Jones, M. A., Stevens, M. P., Logan, S. M., and Wren, B. W. (2009) *Campylobacter jejuni* glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens, *Infect. Immun. 77*, 2544-2556.

73. Twine, S. M., Paul, C. J., Vinogradov, E., McNally, D. J., Brisson, J. R., Mullen, J. A., McMullin, D. R., Jarrell, H. C., Austin, J. W., Kelly, J. F., and Logan, S. M. (2008) Flagellar glycosylation in *Clostridium botulinum*, *Febs J. 275*, 4428-4444.

74. Tzeng, Y. L., and Stephens, D. S. (2000) Epidemiology and pathogenesis of *Neisseria meningitidis*, *Microbes Infect. 2*, 687-700.

75. Strelow, V. L., and Vidal, J. E. (2013) Invasive meningococcal disease, *Arq. Neuro-Psiquiatr. 71*, 653-658.

76. Feil, E. J., and Spratt, B. G. (2001) Recombination and the population structures of bacterial pathogens, *Annu. Rev. Microbiol. 55*, 561-590.

77. van Deuren, M., Brandtzaeg, P., and van der Meer, J. W. M. (2000) Update on meningococcal disease with emphasis on pathogenesis and clinical management, *Clin. Microbiol. Rev. 13*, 144-+.

78. Irving, T. J., Blyuss, K. B., Colijn, C., and Trotter, C. L. (2012) Modelling meningococcal meningitis in the African meningitis belt, *Epidemiol. Infect. 140*, 897-905.

79. Jafri, R. Z., Ali, A., Messonnier, N. E., Tevi-Benissan, C., Durrheim, D., Eskola, J., Fermon, F., Klugman, K. P., Ramsay, M., Sow, S., Shao, Z. J., Bhutta, Z. A., and Abramson, J. (2013) Global epidemiology of invasive meningococcal disease, *Popul. Health Metr. 11*.

80. Uiterwijk, A., and Koehler, P. J. (2012) A history of acute bacterial meningitis, *J. Hist. Neurosci. 21*, 293-313.

81. Stephens, D. S., Greenwood, B., and Brandtzaeg, P. (2007) Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*, *The Lancet 369*, 2196-2210.

82. Unkmeir, A., Kammerer, U., Stade, A., Hubner, C., Haller, S., Kolb-Maurer, A., Frosch, M., and Dietrich, G. (2002) Lipooligosaccharide and polysaccharide capsule: Virulence factors of *Neisseria meningitidis* that determine meningococcal interaction with human dendritic cells, *Infect. Immun. 70*, 2454-2462.

83. Merz, A. J., and So, M. (2000) Interactions of pathogenic *Neisseriae* with epithelial cell membranes, *Annu. Rev. Cell Dev. Biol. 16*, 423-+.

84. Heyderman, R. S., and British Infection, S. (2005) Early management of suspected bacterial meningitis and meningococcal septicaemia in immunocompetent adults - second edition, *J. Infect. 50*, 373-374.

85. Chaudhuri, A., Martin, P. M., Kennedy, P. G. E., Seaton, R. A., Portegies, P., Bojar, M., Steiner, I., and Force, E. T. (2008) EFNS guideline on the management of community-acquired bacterial meningitis: report of an EFNS Task Force on acute bacterial meningitis in older children and adults, *Eur. J. Neurol. 15*, 649-659.

86. Karou, S. D., Balaka, A., Bamoke, M., Tchelougou, D., Assih, M., Anani, K., Agbonoko, K., Simpore, J., and de Souza, C. (2012) Epidemiology and antibiotic resistance of bacterial meningitis in Dapaong, northern Togo, *Asian Pac. J. Trop. Med. 5*, 848-852.

87. Levin, R. E. (2007) *Campylobacter jejuni*: A review of its characteristics, pathogenicity, ecology, distribution, subspecies characterization and molecular methods of detection, *Food Biotechnol. 21*, 271-347.

88. Linton, D., Karlyshev, A. V., Hitchen, P. G., Morris, H. R., Dell, A., Gregson, N. A., and Wren, B. W. (2000) Multiple *N*-acetylneuraminic acid synthetase (*neuB*) genes in *Campylobacter jejuni*: Identification and characterization of the gene involved in sialylation of lipo-oligosaccharide, *Mol. Microbiol. 35*, 1120-1134.

89. Chou, W. K., Dick, S., Wakarchuk, W. W., and Tanner, M. E. (2005) Identification and characterization of *NeuB3* from *Campylobacter jejuni* as a pseudaminic acid synthase, *J. Biol. Chem. 280*, 35922-35928.

90. Sundaram, A. K., Pitts, L., Muhammad, K., Wu, J., Betenbaugh, M., Woodard, R. W., and Vann, W. F. (2004) Characterization of *N*-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*, *Biochem. J. 383*, 83-89.

91. Humphrey, T., O'Brien, S., and Madsen, M. (2007) *Campylobacters* as zoonotic pathogens: A food production perspective, *Int. J. Food Microbiol. 117*, 237-257.

92. Dasti, J. I., Tareen, A. M., Lugert, R., Zautner, A. E., and Groß, U. (2010) *Campylobacter jejuni*: A brief overview on pathogenicity-associated factors and disease-mediating mechanisms, *Int. J. Med. Microbio. 300*, 205-211.

93. Adak, G. K., Long, S. M., and O'Brien, S. J. (2002) Trends in indigenous foodborne disease and deaths, England and Wales: 1992 to 2000, *Gut 51*, 832-841.

94. Helms, M., Simonsen, J., Olsen, K. E. P., and Mølbak, K. (2005) Adverse health events associated with antimicrobial drug resistance in *Campylobacter* species: A registry-based cohort study, *J.Infect. Dis. 191*, 1050-1055.

95. Hwang, T. S., Hung, C. H., Teo, C. F., Chen, G. T., Chang, L. S., Chen, S. F., Chen, Y. J., and Lin, C. H. (2002) Structural characterization of *Escherichia coli* sialic acid synthase, *Biochem. Biophys. Res. Commun. 295*, 167-173.

96. Chen, H., Blume, A., Zimmermann-Kordmann, M., Reutter, W., and Hinderlich, S. (2002) Purification and characterization of *N*-acetylneuraminic acid-9-phosphate synthase from rat liver, *Glycobiology 12*, 65-71.

97. Murkin, A. S., Chou, W. K., Wakarchuk, W. W., and Tanner, M. E. (2004) Identification and mechanism of a bacterial hydrolyzing UDP-*N*-acetylglucosamine 2-epimerase, *Biochemistry 43*, 14290-14298.

98. Milewski, S., Gabriel, L., and Olchowy, J. (2006) Enzymes of UDP-GlcNAc biosynthesis in yeast, *Yeast 23*, 1-14.

99. Samuel, J., and Tanner, M. E. (2004) Active site mutants of the "non-hydrolyzing" UDP-N-acetylglucosamine 2-epimerase from *Escherichia coli*, *BBA-Proteins Proteomics 1700*, 85-91.

100. Morgan, P. M., Sala, R. F., and Tanner, M. E. (1997) Eliminations in the reactions catalyzed by UDP-*N*-acetylglucosamine 2-epimerase, *J. Am. Chem. Soc. 119*, 10269-10277.

101. Campbell, R. E., Mosimann, S. C., Tanner, M. E., and Strynadka, N. C. J. (2000) The structure of UDP-*N*-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases, *Biochemistry 39*, 14993-15001.

102. Velloso, L. M., Bhaskaran, S. S., Schuch, R., Fischetti, V. A., and Stebbins, C. E. (2008) A structural basis for the allosteric regulation of non-hydrolysing UDP-GlcNAc 2-epimerases, *EMBO Rep. 9*, 199-205.

103. Hao, J. J., Balagurumoorthy, P., Suryakala, S., and Sundaramoorthy, M. (2005) Cloning, expression, and characterization of sialic acid synthases, *Biochemical and Biophysical Research Communications 338*, 1507-1514.

104. Blacklow, R. S., and Warren, L. (1962) Biosynthesis of sialic acids by *Neisseria meningitidis J. Biol. Chem. 237*, 3520-&.

105. Suryanti, V., Nelson, A., and Berry, A. (2003) Cloning, over-expression, purification, and characterisation of *N*-acetylneuraminate synthase from *Streptococcus agalactiae*, *Protein Expr. Purif. 27*, 346-356.

106. Li, Y. H., Yu, H., Cao, H. Z., Muthana, S., and Chen, X. (2012) *Pasteurella multocida* CMP-sialic acid synthetase and mutants of *Neisseria meningitidis* CMP-sialic acid synthetase with improved substrate promiscuity, *Appl. Microbiol. Biotechnol. 93*, 2411-2423.

107. Vimr, E. R., and Troy, F. A. (1985) Identification of an inducible catabolic system for sialic acids in *Escherichia coli*, *J. Bact. 164*, 845-853.

108. Krause, S., Hinderlich, S., Amsili, S., Horstkorte, R., Wiendl, H., Argov, Z., Mitrani-Rosenbaum, S., and Lochmuller, H. (2005) Localization of UDP-GlcNAc 2-epimerase/ManAc kinase (GNE) in the Golgi complex and the nucleus of mammalian cells, *Exp. Cell Res. 304*, 365-379.

109. Benie, A. J., Blume, A., Schmidt, R. R., Reutter, W., Hinderlich, S., and Peters, T. (2004) Characterization of ligand binding to the bifunctional key enzyme in the sialic acid biosynthesis by NMR - II. Investigation of the ManNAc kinase functionality, *J. Biol. Chem. 279*, 55722-55727.

110. Blume, A., Ghaderi, D., Liebich, V., Hinderlich, S., Donner, P., Reutter, W., and Lucka, L. (2004) UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase, functionally expressed in and purified from *Escherichia coli*, yeast, and insect cells, *Protein Express. Purif. 35*, 387-396.

111. Munster-Kuhnel, A. K., Tiralongo, J., Krapp, S., Weinhold, B., Ritz-Sedlacek, V., Jacob, U., and Gerardy-Schahn, R. (2004) Structure and function of vertebrate CMP-sialic acid synthetases, *Glycobiology 14*, 43R-51R.

112. Gunawan, J., Simard, D., Gilbert, M., Lovering, A. L., Wakarchuk, W. W., Tanner, M. E., and Strynadka, N. C. J. (2005) Structural and mechanistic analysis of sialic acid synthase *NeuB* from *Neisseria meningitidis* in complex with Mn^{2+} phosphoenolpyruvate, and *N*-acetylmannosaminitol, *J. Biol. Chem. 280*, 3555-3563.

113. Liu, F., Lee, H. J., Strynadka, N. C. J., and Tanner, M. E. (2009) Inhibition of *Neisseria meningitidis* sialic acid synthase by a tetrahedral intermediate analogue, *Biochemistry 48*, 9194-9201.

114. Joseph, D. D. A., Jiao, W. T., and Parker, E. J. (2013) Arg314 is essential for catalysis by *N*-acetyl neuraminic acid synthase from *Neisseria meningitidis*, *Biochemistry 52*, 2609-2619.

115. Hamada, T., Ito, Y., Abe, T., Hayashi, F., Güntert, P., Inoue, M., Kigawa, T., Terada, T., Shirouzu, M., Yoshida, M., Tanaka, A., Sugano, S., Yokoyama, S., and Hirota, H. (2006) Solution structure of the antifreeze-like domain of human sialic acid synthase, *Prot. sci. Pub. Prot. Soc. 15*, 1010-1016.

116. Kristiansen, E., and Zachariassen, K. E. (2005) The mechanism by which fish antifreeze proteins cause thermal hysteresis, *Cryobiology 51*, 262-280.

117. Baardsnes, J., and Davies, P. L. (2001) Sialic acid synthase: the origin of fish type III antifreeze protein?, *Trends Biochem.Sci. 26*, 468-469.

118. Dotson, G. D., Dua, R. K., Clemens, J. C., Wooten, E. W., and Woodard, R. W. (1995) Overproduction and one-step purification of *Escerichia coli* 3-deoxy-D-manno-octulosonic acid 8-phosphate synthase and oxygen-transfer studies during catalysis using isotopic-shifted heteronuclear NMR, *J. Biol. Chem. 270*, 13698-13705.

119. Deleo, A. B., and Sprinson, D. B. (1968) Mechanism of 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthetase, *Biochem. Biophys. Res. Comm. 32*, 873-&.

120. El Zoeiby, A., Sanschagrin, F., and Levesque, R. C. (2003) Structure and function of the Mur enzymes: development of novel inhibitors, *Mol. Microbiol. 47*, 1-12.

121. Anderson, K. S., Sammons, R. D., Leo, G. C., Sikorski, J. A., Benesi, A. J., and Johnson, K. A. (1990) Observation by C-13 NMR of the EPSP synthase tetrahedral intermediate bound to the enzyme active-site, *Biochemistry 29*, 1460-1465.

122. Bravo, I. G., Garcia-Vallve, S., Romeu, A., and Reglero, A. (2004) Prokaryotic origin of cytidylyltransferases and alpha-ketoacid synthases, *Trends Microbiol. 12*, 120-128.

123. Walker, S. R., and Parker, E. J. (2006) Synthesis and evaluation of a mechanism-based inhibitor of a 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase, *Bioorg. Med. Chem. Lett. 16*, 2951-2954.

124. Schofield, L. R., Anderson, B. F., Patchett, M. L., Norris, G. E., Jameson, G. B., and Parker, E. J. (2005) Substrate ambiguity and crystal structure of Pyrococcus furiosus 3-Deoxy-D-arabino-heptulosonate-7-phosphate synthase: An ancestral 3-deoxyald-2-ulosonate-phosphate synthase?, *Biochemistry 44*, 11950-11962.

125. Webby, C. J., Baker, H. M., Lott, J. S., Baker, E. N., and Parker, E. J. (2005) The structure of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from Mycobacterium tuberculosis reveals a common catalytic scaffold and ancestry for type I and type II enzymes, *J. Mol. Biol. 354*, 927-939.

126. Bentley, R., and Haslam, E. (1990) The shikimate pathway - A metabolic tree with many branches, *Crit. Rev. Biochem. Mol. Bio. 25*, 307-384.

127. Cross, P. J., and Parker, E. J. (2013) Allosteric inhibitor specificity of *Thermotoga maritima* 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, *FEBS Letters 587*, 3063-3068.

128. Nelson, S. K., Kelleher, A., Robinson, G., Reiling, S., and Asojo, O. A. (2013) Structure of 2-keto-3-deoxy-D-manno-octulosonate-8-phosphate synthase from *Pseudomonas aeruginosa*, *Acta Crystallogr. F-Struct. Biol. Cryst. Commun. 69*, 1084-1088.

129. Webby, C. J., Baker, H. M., Lott, J. S., Baker, E. N., and Parker, E. J. (2005) The structure of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Mycobacterium tuberculosis* reveals a common catalytic scaffold and ancestry for type I and type II enzymes, *J. Mol. Bio. 354*, 927-939.

130. Radaev, S., Dastidar, P., Patel, M., Woodard, R. W., and Gatti, D. L. (2000) Structure and mechanism of 3-deoxy-D-manno-octulosonate 8-phosphate synthase, *J. Biol. Chem. 275*, 9476-9484.

131. Rick, P. D., and Young, D. A. (1982) Isolation and characterisation of a temperature-sensitive lethal mutant of *Salmonella typherium* that is conditionally deffective in 3-deoxy-D-manno-octuolosonate 8-phosphate synthesis, *J. Bact. 150*, 447-455.

132. Rick, P. D., Neumeyer, B. A., and Young, D. A. (1984) Effect of altered lipid-A synthesis on the synthesis of major proteins of the *Salmonella typhimurium* outer-membrane, *Rev. Infect. Dis. 6*, 455-462.

133. Tao, P., Schlegel, H. B., and Gatti, D. L. (2010) Common basis for the mechanism of metallo and non-metallo KDO8P synthases, *J. Inorg. Biochem. 104*, 1267-1275.

134. Lewis, A. L., Desa, N., Hansen, E. E., Knirel, Y. A., Gordon, J. I., Gagneux, P., Nizet, V., and Varki, A. (2009) Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure, *Proc. Natl. Acad. Sci. U. S. A. 106*, 13552-13557.

135. Schoenhofen, I. C., Vinogradov, E., Whitfield, D. M., Brisson, J. R., and Logan, S. M. (2009) The CMP-legionaminic acid pathway in *Campylobacter*: Biosynthesis involving novel GDP-linked precursors, *Glycobiology 19*, 715-725.

136. Luneberg, E., Zetzmann, N., Alber, D., Knirel, Y. A., Kooistra, O., Zahringer, U., and Frosch, M. (2000) Cloning and functional characterization of a 30 kb gene locus required for lipopolysaccharide biosynthesis in *Legionella pneumophila*, *Int. J. Med. Microbio. 290*, 37-49.

137. Rangarajan, E. S., Proteau, A., Cui, Q. Z., Logan, S. M., Potetinova, Z., Whitfield, D., Purisima, E. O., Cygler, M., Matte, A., Sulea, T., and Schoenhofen, I. C. (2009) Structural and functional analysis of *Campylobacter jejuni* PseG: A UDP-sugar hydrolase from the pseudaminic acid biosynthetic pathway, *J. Biol. Chem. 284*, 20989-21000.

138. Liu, F., and Tanner, M. E. (2006) *PseG* of pseudaminic acid biosynthesis - A UDP-sugar hydrolase as a masked glycosyltransferase, *J. Biol. Chem. 281*, 20902-20909.

139. Lin, B. X., Zhang, Z. J., Liu, W. F., Dong, Z. Y., and Tao, Y. (2013) Enhanced production of *N*-acetyl-D-neuraminic acid by multi-approach whole-cell biocatalyst, *App. Microbio. Biotech. 97*, 4775-4784.

140. Groher, A., and Hoelsch, K. (2012) Mechanistic model for the synthesis of *N*-acetylneuraminic acid using *N*-acetylneuraminate lyase from *Escherichia coli* K12, *J. Mol. Catal. B-Enzym. 83*, 1-7.

141. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Res. 31*, 3784-3788.

142. Louis-Jeune, C., Andrade-Navarro, M. A., and Perez-Iratxeta, C. (2012) Prediction of protein secondary structure from circular dichroism using theoretically derived spectra, *Prot. Struct. Funct. Bioinfo. 80*, 374-381.

143. Marianayagam, N. J., Sunde, M., and Matthews, J. M. (2004) The power of two: Protein dimerization in biology, *Trends Biochem.Sci. 29*, 618-625.

144. Gronenborn, A. M. (2009) Protein acrobatics in pairs; dimerization via domain swapping, *Curr. Opin. Struct. Bio. 19*, 39-49.

145. Ostermeier, M., and Benkovic, S. J. (2001) Evolution of protein function by domain swapping, *Adv. Prot. Chem. 55*, 29-77.

146. Liu, Y., and Eisenberg, D. (2002) 3D domain swapping: As domains continue to swap, *Protein Sci. 11*, 1285-1299.

147. Nishitani, Y., Aono, R., Nakamura, A., Sato, T., Atomi, H., Imanaka, T., and Miki, K. (2013) Structure analysis of archaeal AMP phosphorylase reveals two unique modes of dimerisation, *J. Mol. Biol. 425*, 2709-2721.

148. Froese, D. S., Krojer, T., Wu, X. C., Shrestha, R., Kiyani, W., von Delft, F., Gravel, R. A., Oppermann, U., and Yue, W. W. (2012) Structure of MMACHC reveals an arginine-rich pocket and a domain-swapped dimer for Its B-12 processing function, *Biochemistry 51*, 5083-5090.

149. Alm, M., Pietersma, A. L., Schofield, L. R., and Parker, E. J. (2005) Mechanistic divergence of two closely related aldol-like enzyme-catalysed reactions, *Org. Biomol. Chem. 3*, 4046-4049.

150. Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J., and Schubert, D. (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: Strategies and application to model systems, *Biophys. J. 82*, 1096-1111.

151. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling, *Biophys. J. 78*, 1606-1619.

152. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

153. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates, *J. Appl. Crystallogr. 28*, 768-773.

154. Fischer, H., Neto, M. O., Napolitano, H. B., Craievich, A. F., and Polikarpov, I. (2010) The molecular weight of proteins in solution can be determined from a single SAXS measurement on a relative scale, *J. App. Crystal.*, 101-109.

155. Larkin MA, B. G., Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG. (2007) ClustalW and ClustalX version 2, *Bioinformatics 23*, 2947-2948.

156. Goujon M, M. H., Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI, *Nuc. acids res. 38* 695-699.

157. Rambaut, A. http://tree.bio.ed.ac.uk/software/figtree/.

158. Vann, W. F., Tavarez, J. J., Crowley, J., Vimr, E., and Silver, R. P. (1997) Purification and characterization of the *Escherichia coli* K1 *neuB* gene product *N*-acetylneuraminic acid synthetase, *Glycobiology 7*, 697-701.

159. Waterhouse, A. M. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench, *Bioinformatics doi*.

160. Wahlund, T. M., Woese, C. R., Castenholz, R. W., and Madigan, M. T. (1991) A thermophilic green sulfur bacterium from New Zealand hot springs, *Chlorobium tepidum Arch. Microbiol. 156*, 81-90.

161. Liav, A. L., and Sharon, N. (1973) Synthesis of 2,4-diacetamido-2,4,6-trideoxy-L-altrose, -L-idose and -L-talose from benzyl 6-deoxy-3,4-*O*-isopropylidene-beta-L-galactopyranoside, *Carb. Res. 30*, 109-126.

162. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics 22*, 195-201.

163. Benkert, P., Biasini, M., and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models, *Bioinformatics 27*, 343-350.

164. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Bio. 215*, 403-410.

165. Williams, A. E., and Blakemore, W. F. (1990) Pathogenesis of meningitis caused by *Streptococcus suis* type 2, *J.Infect. Dis. 162*, 474-481.

166. Robertson, I. D., and Blackmore, D. K. (1990) Experimental studies on the comparative infectivity and pathogenicity of *Streptococcus suis* type 2: Porcine and human isolates in pigs, *Epidemiol. Infect. 105*, 469-478.

167. Staats, J. J., Feder, I., Okwumabua, O., and Chengappa, M. M. (1997) *Streptococcus suis*: Past and present, *Vet. Res. Commun. 21*, 381-407.

168. Wang, K. C., Zhang, W., Li, X. C., Lu, C. P., Chen, J. M., Fan, W. X., and Huang, B. X. (2013) Characterization of *Streptococcus suis* isolates from slaughter swine, *Curr. Microbiol. 66*, 344-349.

169. Hanterdsith, B., Tharavichitkul, P., Mahanupab, P., and Raksamat, W. (2013) Postmortem diagnosis of sudden unexpected death from *Streptococcus suis* type 2 infection: A case report, *J. Forensic Leg. Med. 20*, 347-349.

170. Zalas-Wiecek, P., Michalska, A., Grabczewska, E., Olczak, A., Pawlowska, M., and Gospodarek, E. (2013) Human meningitis caused by *Streptococcus suis*, *J. Med. Microbiol. 62*, 483-485.

171. Arends, J. P., and Zanen, H. C. (1988) Meningitis caused by *Streptococcus suis* in humans, *Rev. Infect. Dis. 10*, 131-137.

172. Durand, F., Perino, C. L., Recule, C., Brion, J. P., Kobish, M., Guerber, F., and Croize, J. (2001) Bacteriological diagnosis of *Streptococcus suis* meningitis, *Eur. J. Clin. Microbiol. Infect. Dis. 20*, 519-521.

173. Cowan, J. A. (2002) Structural and catalytic chemistry of magnesium-dependent enzymes, *Biometals 15*, 225-235.

174. Matoba, Y., Tanaka, N., Noda, M., Higashikawa, F., Kumagai, T., and Sugiyama, M. (2013) Crystallographic and mutational analyses of tannase from *Lactobacillus plantarum*, *Prot. Struct. Funct. Bioinfo. 81*, 2052-2058.

175. Tsitsanou, K. E., Oikonomakos, N. G., Zographos, S. E., Skamnaki, V. T., Gregoriou, M., Watson, K. A., Johnson, L. N., and Fleet, G. W. J. (1999) Effects of commonly used cryoprotectants on glycogen phosphorylase activity and structure, *Prot. Sci. 8*, 741-749.

176. Johal, A. R., Schuman, B., Alfaro, J. A., Borisova, S., Seto, N. O. L., and Evans, S. V. (2012) Sequence-dependent effects of cryoprotectants on the active sites of the human ABO(H) blood group A and B glycosyltransferases, *Acta Crystallogr. Sect. D-Biol. Crystallogr. 68*, 268-276.

177. Finne, J. (1985) Polysialic acid - a glycoprotein carbohydrate involved in neural adhesion and bacterial-meningitis, *Trends Biochem.Sci. 10*, 129-132.

178. Varki, A. (1997) Sialic acids as ligands in recognition phenomena, *FASEB Journal 11*, 248-255.

179. Edwards, U., Muller, A., Hammerschmidt, S., Gerardy-Schahn, R., and Frosch, M. (1994) Molecular analysis of the biosynthesis pathway of the α-2,8 polysialic acid capsule by *Neisseria meningitidis* serogroup B, *Mol. Microbiol. 14*, 141-149.

180. Preston, A., Mandrell, R. E., Gibson, B. W., and Apicella, M. A. (1996) The lipooligosaccharides of pathogenic gram-negative bacteria, *Crit. Rev. Microbiol. 22*, 139-180.

181. Mizanur, R. M., and Pohl, N. L. (2008) Bacterial CMP-sialic acid synthetases: production, properties, and applications, *Appl. Microbiol. Biotechnol. 80*, 757-765.

182. Hao, J., Vann, W. F., Hinderlich, S., and Sundaramoorthy, M. (2006) Elimination of 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid 9-phosphate synthase activity from human *N*-acetylneuraminic acid 9-phosphate synthase by a single mutation, *Biochem. J. 397*, 195-201.

183. Cross, P. J., Dobson, R. C. J., Patchett, M. L., and Parker, E. J. (2011) Tyrosine latching of a regulatory gate affords allosteric control of aromatic amino acid biosynthesis, *J. Biol. Chem. 286*, 10216-10224.

184. Chen, C. K. M., Lee, G. C., Ko, T. P., Guo, R. T., Huang, L. M., Liu, H. J., Ho, Y. F., Shaw, J. F., and Wang, A. H. J. (2009) Structure of the alkalohyperthermophilic *Archaeoglobus fulgidus* lipase contains a unique C-terminal domain essential for long-chain substrate binding, *J. Mol. Biol. 390*, 672-685.

185. Kim, J. S., Cha, S. S., Kim, H. J., Kim, T. J., Ha, N. C., Oh, S. T., Cho, H. S., Cho, M. J., Kim, M. J., Lee, H. S., Kim, J. W., Choi, K. Y., Park, K. H., and Oh, B. H. (1999) Crystal structure of a maltogenic amylase provides insights into a catalytic versatility, *J. Biol. Chem. 274*, 26279-26286.

186. Vanoni, M. A., and Curti, B. (2008) Structure-function studies of glutamate synthases: A class of self-regulated iron-sulfur flavoenzymes essential for nitrogen assimilation, *Iubmb Life 60*, 287-300.

187. Reaves, M. L., Lopez, L. C., and Daskalova, S. M. (2008) Replacement of the antifreeze-like domain of human *N*-acetylneuraminic acid phosphate synthase with the mouse antifreeze-like domain impacts both *N*-acetylneuraminic acid 9-phosphate synthase and 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid 9-phosphate synthase activities, *BMB Rep. 41*, 72-78.

188. Bezerra, G. A., Dobrovetsky, E., Viertlmayr, R., Dong, A., Binter, A., Abramic, M., Macheroux, P., Dhe-Paganon, S., and Gruber, K. (2012) Entropy-driven binding of opioid peptides induces a large domain motion in human dipeptidyl peptidase III, *Proc. Natl. Acad. Sci. U. S. A. 109*, 6525-6530.

189. Biela, A., Sielaff, F., Terwesten, F., Heine, A., Steinmetzer, T., and Klebe, G. (2012) Ligand binding stepwise disrupts water network in thrombin: enthalpic and entropic changes reveal classical hydrophobic effect, *J. Med. Chem. 55*, 6094-6110.

190. Harrison, A. N., Reichau, S., and Parker, E. J. (2012) Synthesis and evaluation of tetrahedral intermediate mimic inhibitors of 3-deoxy-D-manno-octulosonate 8-phosphate synthase, *Bioorg. Med. Chem. L. 22*, 907-911.

191. Walker, S. R., Cumming, H., and Parker, E. J. (2009) Substrate and reaction intermediate mimics as inhibitors of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, *Org. Biomol. Chem. 7*, 3031-3035.

192. Schrödinger, L. (2011) The PyMOL molecular graphics system, version 1.5.

193. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol. 302*, 205-217.

194. Hanahan, D. (1983) STUDIES ON TRANSFORMATION OF ESCHERICHIA-COLI WITH PLASMIDS, *J. Mol. Biol. 166*, 557-580.

195. Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005) Protein Identification and Analysis Tools on the ExPASy Server, *John M. Walker (ed): The Proteomics Protocols Handbook*, 571-607

196. Schoner, R., and Herrmann, K. M. (1976) 3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase - Purification, properties, and kinetics of tyrosine-sensitive isoenzyme from *Escherichia coli J. Biol. Chem. 251*, 5440-5447.

197. Leatherbarrow, R. J. (2001) GraFit Version 5, 5.0 ed., Erithacus Software Limited, Horley, UK.

198. Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies, *Analytical Biochemistry 357*, 289-298.

199. Laue, T. M., Shah, D. B., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In analytical ultracentrifugation in biochemistry and polymer science *T. Roy. Soc. Chem. Camb.*, 90–125.

200. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Crystallogr. 25*, 495-503.

201. Mills, R. D., Trewhella, J., Qiu, T. W., Welte, T., Ryan, T. M., Hanley, T., Knott, R. B., Lithgow, T., and Mulhern, T. D. (2009) Domain organization of the monomeric form of the Tom70 mitochondrial import receptor, *J. Mol. Biol. 388*, 1043-1058.

202. McPhillips, T. M., McPhillips, S. E., Chiu, H. J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M., and Kuhn, P. (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines, *J. Synchrot. Radiat. 9*, 401-406.

203. Bailey, S. (1994) The CCP4 suite - programs for protein crystallography *Acta Crystallogr. D Biol. Crystallogr. 50*, 760-763.

204. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallographica Section D-Biological Crystallography 53*, 240-255.

205. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D-Biological Crystallography 60*, 2126-2132.

206. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallographica Section D-Biological Crystallography 66*, 12-21.

207. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem. 47*, 1739-1749.

208. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes, *J. Med. Chem. 49*, 6177-6196.

209. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *J. Med. Chem. 47*, 1750-1759.

210. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comp. Chem. 26*, 1781-1802.

211. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graph. Model. 14*, 33-38.

212. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B. 102*, 3586-3616.

213. KEGG. (2013) Kyoto encyclopaedia of genes and genomes http://www.genome.jp/kegg/